

---

## **Regional Policy**



The MIT Press  
Cambridge, Massachusetts, and London, England

---

**Regional Policy**  
**Readings in Theory and Applications**

Edited by John Friedmann  
and William Alonso

711.3  
F9112  
Cop. 2

Copyright © 1975 by  
The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

This book was printed on Finch Title 93  
and bound in MBV-4422  
by The Colonial Press Inc.  
in the United States of America

**Library of Congress Cataloging in Publication Data**

Friedmann, John, comp.  
Regional policy.

Edition for 1964 published under title:  
Regional development and planning.

Bibliography: p.

1. Regional planning—Addresses, essays, lectures.

I. Alonso, William, joint comp. II. Title.

HT 391. F75 1974 309.2'5 74-7412

ISBN 0-262-06057-4

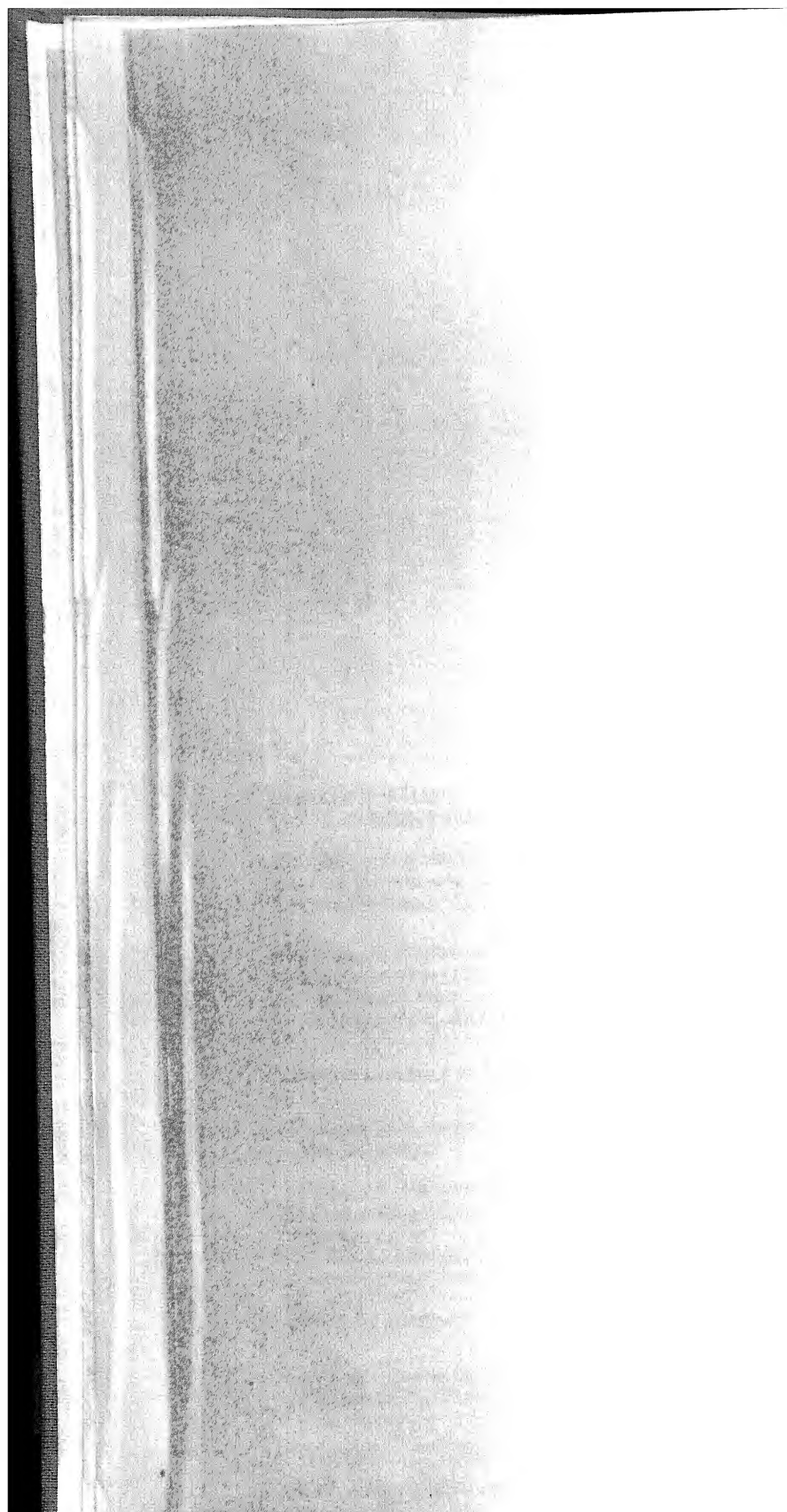
This reader represents a major revision of *Regional Development and Planning: A Reader*, edited by John Friedmann and William Alonso, published by The MIT Press in 1964.

---

For Walter Isard

DEC 01 '77

HUNT LIBRARY  
CARNEGIE-MELLON UNIVERSITY  
PITTSBURGH, PENNSYLVANIA 15213



---

## Contents

Preface xi

Contributors xiii

Introduction xv

### Part I Concepts of Space and Development

Introductory Note 3

#### Location and Spatial Structure

- 1 An Outline of a Stochastic Model for the Study of the Spatial Structure  
and Development of a Human Population System 9

Rutledge Vining

- 2 Location Theory 35

William Alonso

- 3 Industrial Location and Regional Policy in Economic Development 64

William Alonso

- 4 The Nature of Economic Regions 97

August Lösch

- 5 The Geography of the United States in the Year 2000 106

Brian J. L. Berry

#### Spatial Systems in Economic Development

- 6 Interregional and International Transmission of Economic Growth 139

Albert O. Hirschman

- 7 Regional Inequality and the Process of National Development: A  
Description of the Patterns 158

Jeffrey G. Williamson

- 8 Internal and External Factors in the Development of Urban  
Economies 201

Wilbur R. Thompson

viii CONTENTS

- 9 Population Movements and the Shape of Urban Growth: Implications  
for Public Policy 221  
Peter A. Morrison
- 10 Tanzania 1920-63: The Spatial Impress of the Modernization  
Process 244  
Peter R. Gould
- 11 The Spatial Organization of Power in the Development of Urban  
Systems 266  
John Friedmann

**Growth in Subnational Regions**

- 12 Natural Resource Endowment and Regional Economic Growth 307  
Harvey Perloff and Lowdon Wingo, Jr.
- 13 Location Theory and Regional Economic Growth 332  
Douglass C. North
- 14 Exports and Regional Economic Growth 348  
Charles M. Tiebout  
A Reply  
Douglass C. North  
Rejoinder  
Charles M. Tiebout
- 15 Industrialization, Factor Markets, and Agricultural Development 358  
William H. Nicholls

**Part II The Role of Cities in National Development**

- Introductory Note 383
- 16 The Cultural Role of Cities 385  
Robert Redfield and Milton B. Singer
  - 17 The City in Informational Society 406  
Toshio Sanuki
  - 18 The Analysis of "Over-Urbanization" 421  
N. V. Sovani
  - 19 The Economics of Urban Size 434  
William Alonso

- 20 The Urban Poor: Disruption or Political Integration in Third World Cities? 451  
Joan Nelson
- 21 The Absorption of Labor in the Urban Economy: The Case of Developing Countries 473  
John Friedmann and Flora Sullivan

### **Part III Formulating the Issues in Regional Policy**

- Introductory Note 505
- 22 National Growth Policy 507  
John D. Ehrlichman
- 23 The National System of Cities as an Object of Public Policy 516  
Wilbur R. Thompson
- 24 Welfare Reform: The Southern View 534  
Tom Herman
- 25 Growth Poles and Growth Centers in Regional Planning: A Review 539  
D. F. Darwent
- 26 Criteria for a Growth Centre Policy 566  
Niles M. Hansen
- 27 The Territorial Basis of National Transportation Planning 588  
John Friedmann and Barbara Stuckey
- 28 Urban and Regional Imbalances in Economic Development 622  
William Alonso
- 29 Problems, Purposes, and Implicit Policies for a National Strategy of Urbanization 636  
William Alonso

### **Part IV Case Studies of Regional Planning**

- Introductory Note 653
- 30 Evolution of Spatial Organization in the Ottoman Empire and Turkish Republic 655  
Ilhan Tekeli



x CONTENTS

- 31 Principles of Planning and Zionist Ideology: The Israeli Development Town 680  
Elisabeth A. Altman and Betsey R. Rosenbaum
- 32 Administrative Pluralism and Italian Regional Planning 695  
Robert C. Fried
- 33 Regional Development Policy in Spain 712  
H. W. Richardson
- 34 The Politics of French Regional Planning 727  
George W. Ross and Stephen S. Cohen
- 35 Methods of Achieving Consistency between National and Regional Location Plans 751  
D. Schejbal and O. Žurek
- 36 Aspects of Regional Planning and Theory in the United States 773  
William Alonso

**Part V Bibliographical Essay**

- 37 Regional Development Planning: The Progress of a Decade 791  
John Friedmann

---

## Preface

Ten years ago we edited a volume titled *Regional Development and Planning* (MIT Press, 1964), which met a most gratifying reception and, after several printings, is still widely used. It served, we believe, to introduce to many the contours of a new field of intellectual inquiry and of policy analysis and action. Since then, regional development planning has become a visible part of the intellectual landscape and a prominent part of national policy in many countries. The resulting profusion of scholarly interest, together with the growing body of policy experience, has deepened, spread, and, in some cases, transformed our understanding of the processes underlying regional change and our perception of the purposes of regional policy. Hence we have decided to offer a new collection of readings.

We have retained only a few readings from the earlier volume, and these primarily for reasons of historical interest. Students new to the study of regional planning should know the intellectual foundations of the field, and the chapters in question have entered the rank of "classics." Among the new readings, three were prepared especially for this volume, including one which gives an overview of the literature published in the last ten years. The remaining essays were chosen because they illustrate in ways we think particularly effective the richness and diversity of the themes with which regional planners have been concerned.

Despite the size of this volume, we have been forced, by reasons of space, to omit many excellent contributions, often most reluctantly. Time and again, even in the final stages of preparation of this collection, we have come across materials we wished we could include. The field is brimming with an effervescence of fresh ideas, new findings, and attractive expressions of insights. We apologize to readers and authors for such omissions. Perhaps some years from now, when more synthetic work summarizes and ties together common strands, some future collection of readings may be able better to epitomize this broad field. The prospect of a unified magisterial treatment in a single volume by a single author still seems as remote to us as it did ten years ago.

J.F.

W.A

October 1974

\_\_\_\_\_

12-11-11

\_\_\_\_\_

12-11-11

\_\_\_\_\_

## Contributors

WILLIAM ALONSO, Professor of Regional Planning, University of California, Berkeley

ELISABETH A. ALTMAN, Massachusetts Department of Mental Health

BRIAN J. L. BERRY, Irving B. Harris Professor of Urban Geography and Director of Training Programs, Center for Urban Studies, University of Chicago

STEPHEN S. COHEN, Associate Professor, Department of City and Regional Planning, University of California, Berkeley

D. F. DARWENT, formerly Research Associate, Institute of Urban and Regional Development, University of California, Berkeley

JOHN D. EHRLICHMAN, formerly Domestic Affairs Adviser to the President of the United States

ROBERT C. FRIED, Professor of Political Science, University of California, Los Angeles

JOHN FRIEDMANN, Professor and Head, Urban Planning Program, School of Architecture and Urban Planning, University of California, Los Angeles

PETER R. GOULD, Professor of Geography, Pennsylvania State University, State College

NILES M. HANSEN, Professor of Economics, University of Texas

TOM HERMAN, Journalist, *The Wall Street Journal*

ALBERT O. HIRSCHMAN, Professor of Economics, Harvard University

AUGUST LÖSCH (deceased), author of *The Economics of Location*

PETER A. MORRISON, Demographer, The Rand Corporation, Santa Monica, California

JOAN NELSON, Research Fellow, The Urban Institute, Washington, D.C.

WILLIAM H. NICHOLLS, Professor of Economics, Vanderbilt University

DOUGLASS C. NORTH, Professor and Chairman, Department of Economics, University of Washington

HARVEY PERLOFF, Professor and Dean, School of Architecture and Urban Planning, University of California, Los Angeles

ROBERT REDFIELD (deceased), formerly Professor of Anthropology, University of Chicago

H. W. RICHARDSON, Professor of Economics and Director, Centre for Research in the Social Sciences, University of Kent

HUNT LIBRARY  
CARNEGIE-MELLON UNIVERSITY  
PITTSBURGH, PENNSYLVANIA 1521

xiv CONTRIBUTORS

BETSEY R. ROSENBAUM, management consultant, New York

GEORGE W. ROSS, Assistant Professor of Sociology, Brandeis University

TOSHIO SANUKI, Lecturer in Economics, Waseda University, Tokyo

D. SCHEJBAL, formerly Head of Department, Ministry for Economic Planning, Prague, Czechoslovakia

MILTON B. SINGER, Professor of Anthropology, University of Chicago

N. V. SOVANI, Director, Gokhale Institute of Politics and Economics, Poona University, Poona, India

BARBARA STUCKEY, doctoral candidate in urban planning, School of Architecture and Urban Planning, University of California, Los Angeles

FLORA SULLIVAN, doctoral candidate in urban planning, School of Architecture and Urban Planning, University of California, Los Angeles

ILHAN TEKELI, Professor of Urban Planning, Faculty of Architecture, Middle East Technical University, Ankara

WILBUR R. THOMPSON, consulting economist, Phoenix, Arizona

CHARLES M. TIEBOUT (deceased), formerly Professor of Economics, University of Washington

RUTLEDGE VINING, Professor of Economics, University of Virginia

JEFFREY G. WILLIAMSON, Professor of Economics, University of Wisconsin

LOWDON WINGO, JR., Professor and Chairman, Department of City and Regional Planning, University of Pennsylvania

O. ŽUREK, formerly of the Ministry for Economic Planning, Prague, Czechoslovakia

## **Introduction**

### **Efficiency, Equity, and National Integration in Regional Planning**

From earliest history, military and political considerations have played a part in the territorial governance of nations and empires, but other goals have come to dominate regional policy in recent years in most countries. Modern concern with regional planning may be said to start with the surprisingly recent social invention of economic development as an explicit national objective, which, for practical purposes, dates to the years immediately following World War II. Then the traditional idea of national prosperity was given concrete meaning and intellectual structure by the macroeconomic concepts which had been developed during the 1920s and 1930s. It soon became apparent, however, that the arithmetic of macroeconomic planning could be made more powerful by the geometry of spatial relationships. The decision about how much of a given resource to allocate to a specific function must ultimately involve the question of where this allocation is to take effect, if only because its contribution to national economic growth will tend to vary greatly with location. Similarly, it soon became clear that certain place-bound complementarities had to be taken into account for reasons of national efficiency: new industry might require related housing, schools, and other facilities; new roads might change agricultural patterns and require new irrigation, storage, and marketing facilities. Because of a myriad of such examples, the pursuit of the goal of national economic growth was broadened to include a regional (or spatial) dimension.

In the early years of regional planning, issues of equity or income distribution were generally ignored. It was assumed that they would be solved by the progressive integration and rationalization of the economy as a function of economic growth, or else they were dismissed as involving only political issues and transfer payments. Only gradually, possibly in the early 1960s, did it come to be recognized that the single criterion of economic growth could not, by itself, represent all there was to the national purpose. Nations were also concerned with inequalities in living conditions among diverse groups, defined by class, occupation, age, race, or other criteria, including residents of certain regions such as rural areas, depressed industrial areas, or urban slums. Economic planners thus began to distinguish between the national goal of economic growth (commonly called the efficiency objective) and the national goal of fairness or equality in the distribution of income or consumption (commonly called the equity objective). Now that two goals were recognized, national policy could not address itself simply

to the maximization of one objective or the other: it had to make trade-offs between them. For, unfortunately, the maximization of growth can frequently worsen the problem of inequality, while the pursuit of equality may be inefficient for national growth.

The matter of equity, however, had gained considerable currency in regional matters even before its recognition by national economic planners, and programs for distressed regions were already firmly established in many countries. Such programs were often independent of or even in conflict with those of the agencies in charge of national planning. For reasons that may have to do with the sociology of professions and institutions, regional planners were more concerned with equity and tended to be architects, geographers, and sociologists, while the growth-oriented national planners were typically economists. Their different vocabularies and professional styles, as well as their different emphasis on objectives, have been a major obstacle to the integration of regional planning with macroeconomic and sectoral planning at the national level, even when national plans have tried to embrace both of these dimensions simultaneously.

Issues of efficiency and equity are profoundly involved in the problem of national integration, both sociocultural and economic. In economic theory, barriers to geographic or sectoral mobility are viewed as imperfections that prevent the system from achieving full efficiency. Economic development in this view may be considered as that integration which eliminates these barriers and imperfections, both increasing total production and, by allowing each person to fulfill his capacities, reducing inequalities. The failure of people to migrate from stagnant to prosperous areas, of capital to seek out pools of unemployed labor, and of new technologies to be adopted simultaneously everywhere, are, in this classical market situation, regarded as imperfections that will correct themselves in the fullness of time or remain minor aberrations in rationality. This classical economic view provides a valuable insight into economic relations, and the process it postulates is visibly taking place in many cases, but its universality is challenged by today's realities on two grounds, one economic and the other social.

In terms of economics, it is quite clear that many of the problems are not disappearing and are in many cases worsening. The reasons why this happens are explored in new theories of perverse and backwash effects, of exploitation and dependency, and they are increasingly clarified by a greater understanding of the dynamics of historical interdependencies. These new approaches have yet to coalesce into a new orthodoxy, but they are a response to the frequent failure of classical economic integration theory.

Similarly, this classical approach to integration has failed to deal with real sociocultural phenomena, largely because it takes an atomistic view of society as composed of individuals. In reality, many countries in Europe, Africa, Asia, and the Americas are multinational societies whose people divide themselves by ethnicity, language, religious belief, and custom and often concentrate in particular regions. These sociocultural groups possess a soli-

darity which escapes the classical analysis but which has profound consequences. In its most dramatic instances, it leads to separatist wars, as in the cases of Bangladesh and Biafra. Even among developed countries, strong separatist movements are frequently found, as among the Quebecois and the Basques, while regional ethnicity is a dominant fact of national life in Great Britain, Yugoslavia, Belgium, the Soviet Union, and many other countries. In the United States, the phenomenon is less visible but is nonetheless present for blacks, Indians, Chicanos, and even for Southerners.

The reality, then, is that there are groups that often play an intermediate role between the individual and the nation. Development and national integration are often in tension with the integrity of these groups, which may feel threatened by the outmigration of their young, by changes in their customs or erosion of their language, or by the arrival of "foreigners" who run government agencies or private enterprises. They often demand greater autonomy or political or economic equality as a group, and are indeed threatened by circumstances or policies that may benefit an individual but separate him from the group. In contrast to the apparent conclusion of the traditional economic view, then, the integration attendant upon economic development may be seen as a threat to the integrity of cultural groups and thereby call forth greater efforts to combat it. Hence, national policy, particularly in its regional dimension, has had to balance its objectives and actions in light of the often conflicting perspectives of individualism and group solidarity.

#### **Integrity of the Environment and Quality of Life**

More recently, two additional goals have become important, especially in the developed countries. These concern the integrity of the environment and the quality of life. Both goals are still poorly articulated and are the subject of at least as much rhetoric as rational analysis, but they are nonetheless real, reflecting the concerns of people and affecting the actions of their governments. Both of them are strongly territorial, and their emergence accounts in part for the resurgence of interest in urban and regional questions.

The goal of environmental integrity spans issues ranging from the possible extinction of the human race to the aesthetics of landscapes and cities. While at one extreme the issues are worldwide rather than local, most problems of the environment are strongly tied to particular locations. This is obviously true in the case of the preservation of scenic and other areas of natural amenity. But environmental issues are generally regional because the environment is a system in which carrying capacities for pollutants and balances of biotic processes are specific to localities. Most of the environmental phenomena that concern us are physical processes that take place over limited, if sometimes large, physical distances. Therefore, policies addressed to the interaction between human activities and these processes necessarily assume a regional cast.

Regard for the quality of life is perhaps the most diffuse but also the most pervasive goal of human societies. Its concerns may include the safety of



people and property in cities; the cultural identity of distinct population groups and collectivities; cultural enrichment; the sense of worth of the individual and his relation to communities; and a variety of other aspects still struggling for articulation. Ultimately, the questions raised are those of the ancient quest for the good life in the good society. These issues have occupied some of the best minds throughout human history, and by comparison much of what is being said on this subject today is naive in its trivialization of the matter. Nonetheless, it is easy to see that this goal calls for a territorial focus to the extent that it deals with individuals in communities whose daily lives and interactions are carried out in particular localities.

These new goals of public policy often conflict with the traditional criteria of efficiency and equity. For instance, many valued aspects of the quality of life, especially among the rich and cultivated classes, are based on the availability of low-cost labor in a variety of crafts and services. Equalization of earnings raises the comparative cost of such things as personalized service in stores and live musical performances and thus changes the texture of daily life. Another frequent conflict arises between environmentalist conservation and the objective of economic growth. Environment and equity often conflict because restrictive conservationist policies tend to limit most sharply the choices of those who have few resources. Even the environmental and quality of life goals are in frequent contradiction, as instanced by our ambivalence with respect to the automobile, which pollutes at the same time that it enlarges the range and freedom of physical movement.

The recognition of the multiplicity of these goals has brought with it a growing awareness that there can be no hope of resolving the dilemmas and trade-offs among them through technical planning alone. Difficult value choices must be made and, further, choices will involve a rather uneven incidence of costs and benefits—hence a growing recognition of the need to proceed through a fusion of technical planning with politics. The very vagueness of the objectives of public policy calls for operational clarification through mutual question and answer between the technical and political realms.

#### **The Knowledge Explosion in Regional Planning**

The gradual development of national goals and their territorial implications has led to a greater experimentation with regional planning approaches than was true a decade ago. It has also resulted in an explosion of research, as the interest in the issues of regional planning has increased and as money and resources have been funneled in this direction. New books, new journals, government and international agency reports, and an important network of mimeographed working papers have added millions of words and some new insights. National and international conferences, both disciplinary and interdisciplinary, sometimes mixing scholars and officials, have helped create some community of understanding, some sharing of experiences, and some diffusion of knowledge. Special research institutes have been created, and

others have reoriented their work in the direction of urban and regional studies. University degree programs specializing in these matters are now common, and many of the traditional disciplines have recognized the urban and regional dimensions as a valid specialty within their field.

While much of this activity has amounted to motion without movement, there has also been substantial progress. Together with that increasing recognition of the vast areas of our ignorance that is said to be the beginning of wisdom, there has been a steady rate of advance in our knowledge and understanding.<sup>1</sup> One has only to look at our introduction to *Regional Development and Planning*, published ten years ago, for confirmation. Many of the things we said there seem very wrong now. We argued, for instance, that regional planning in developed countries would be primarily a mopping-up operation to eliminate certain imperfections and inequities remaining once the integration of the space economy had been achieved in industrially advanced societies. Of course, this has changed radically with the clearer recognition of the stubbornness, if not perversity, of equity issues and of the profound effects of externalities on efficiency. Far from receding in developed countries, regional issues have been gaining in importance, but the substance of the debate has changed.

Three particular developments in regional research deserve to be mentioned. First, urban and regional studies, which had been accorded low status and low priority in the allocation of funds, have gained not only acceptance in most social science disciplines but in some cases also a preferred status. This has given these specialties the inducements of rewards and prestige that have attracted first-rate minds. In consequence, the general quality of studies has improved, and theory and empirical research have become sharper. Many issues have been better delineated, surprises have been uncovered, and questions have been better posed.

A second development is equally significant. Whereas academic research into regional planning questions used to be a virtual American monopoly, excellent centers of relevant scholarship and scholarly associations are found today in many countries of Europe, Latin America, Africa, and Asia. Books and journals focusing on regional and urban development are now being produced in several countries, many of them of excellent quality. Although the United States still dominates in regional theory, the trend is clearly toward an international spread of competence.

The third development to which we wish to call attention is the more mature understanding of the role and limits of formal models and quantitative approaches. While the past few years have seen great advances in theory and technique, in the availability of data, and in the capacity of computers, they have also seen a growing realization that the methods of quantitative analysis provide useful tools but not ready-made answers. There are many reasons

<sup>1</sup>For further detail on advances in our understanding of regional planning issues, the interested reader is referred to chapter 37 of this volume.

for this. We referred earlier to the recognition that the multiplicity of objectives and the distributional consequences of policies require that technical analysis and political processes work in close coordination. On purely technical grounds, there has also been an awakening to such problems as combinatorial complexity, error propagation, dilemmas of aggregation and disaggregation, limitations of functional forms and of statistical techniques, and so forth. As technical intelligence has advanced, it has generated greater understanding of its own limitations.

### **Arrangement of Selections**

In this collection we have tried to include pieces that would provide a survey, however eclectic in disciplinary terms, of what is known about regional development and the experience of regional planning. At the same time, and unavoidably, we have provided the reader with a sampling of the intellectual styles of inquiry and expression in the field. While those who are already specialists in some area will undoubtedly be dissatisfied with our treatment of it, we hope that they will find it interesting to see how other approaches define and deal with related problems. For this reason, we have tried to include essays that are representative of certain approaches, while pursuing at the same time a criterion of choosing what seemed to us good and important pieces. Sometimes, however, these conflicting criteria could not be met, and we have generally opted for excellence over representativeness.

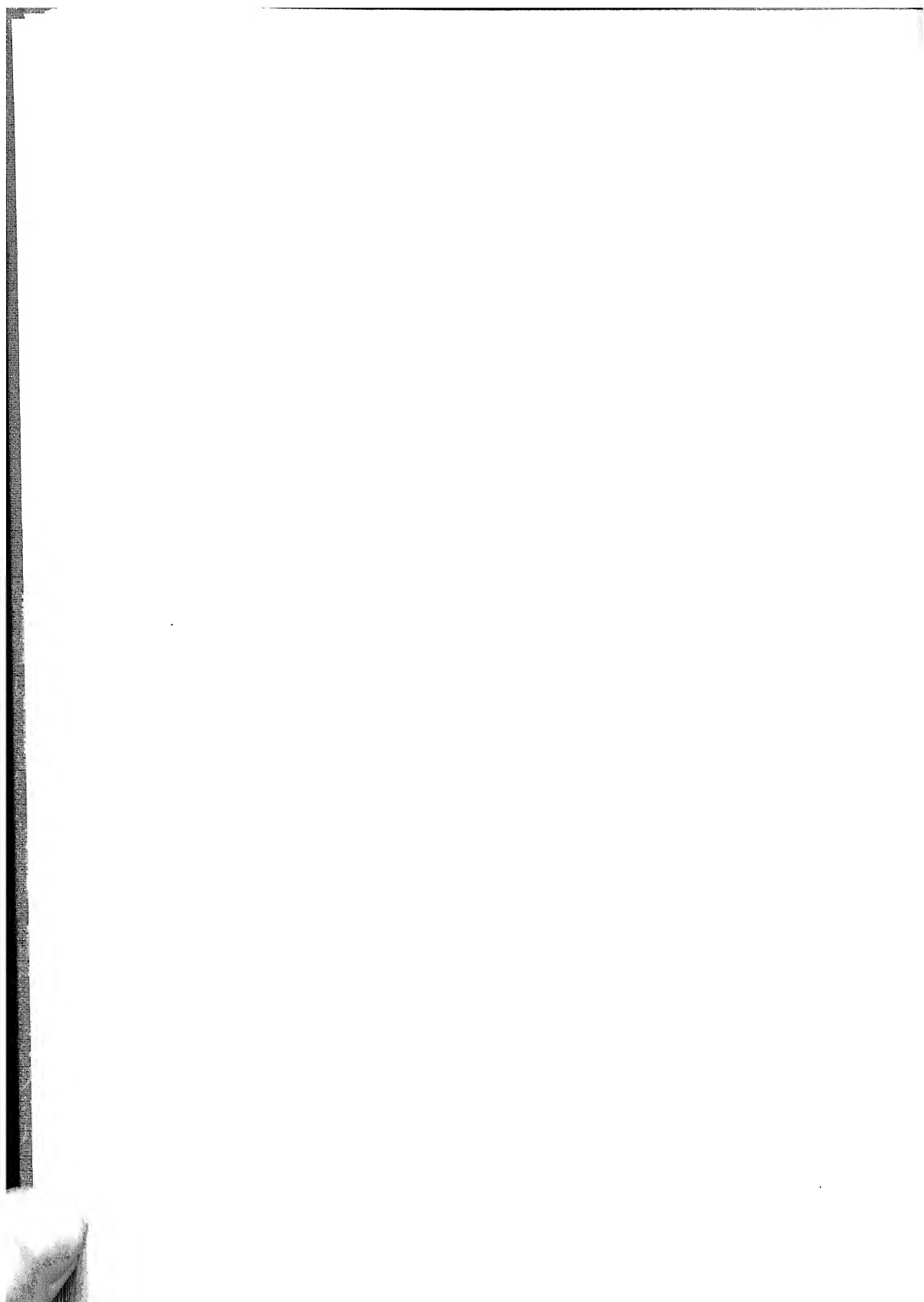
Perhaps the most notable exclusion is that of economics of a highly mathematical or statistical style. Many important contributions have been made by such "hard" approaches of quantitative study, but we decided not to combine them in this single volume with the rather "soft" ones we selected. It is our impression, however, that most of the insights and factual knowledge that the "hard" style has produced are represented in the various essays in a more literary form, and that we are not thereby denying our readers some key insight or information.

The organization of readings as diverse and rich as those in this volume is always arbitrary, and many pieces might have been put in one or another section. Indeed the organization and names of the sections are Procrustean, but there is no help for that. Any other set of groupings would have been equally unsatisfactory.

The readings are grouped into four main parts, each introduced by a brief guide to its essays. Part I, which is the longest, deals with concepts of space and development. It is divided into three sections, focused on location and the spatial structure of nations, on the dynamics of spatial systems development, and on economic growth in subnational regions. Part II is concerned with the special role of cities in national development and with some of the controversies that revolve around the concept of urbanization. Logically, this should have been followed by selections on rural and economically distressed areas. These topics are an important theme in many of the chapters, but we found no way of grouping them together that satisfied us, and thus

they are dealt with throughout the book. Part III deals with the principal issues in regional development policy, while part IV presents case studies of regional planning for both developed and developing countries. In overview, the arrangement presents, first, the pattern of national territorial systems, then the dynamics of the pattern, then issues of policy formulation, and finally some review of recent experience. The liveliness of the material, of course, spills over these categories.

A particular matter which troubled us is the problem of repetition of certain concepts from one piece to another, and we were unable to eliminate it. In our defense we cite two grounds. First, the ideas that keep reappearing are central ones, or at least are so perceived by authors in this field. Since few readers will read the volume through, but rather sample it over time, the overlap helps insure an early acquaintance with central themes in the literature. Second, the attentive reader will note that many such repetitions are often in reality variations, the differences among which are subtle but important because they both reveal the intellectual ancestry and implicit assumptions of the writers, and because they often lead them to different conclusions.



---

**Part I Concepts of Space and Development**



---

## Introductory Note

Human activities are distributed over the national territory in certain rhythms and patterns that are neither arbitrary nor the workings of pure chance. They result rather from the history and interdependencies that give form to socioeconomic space. The social geography of a country is consequently a projection upon the two dimensions of a map of the multidimensional web of connectivities that constitute the social system. The organization and evolution in space of a society can only be understood in this larger context. Attempts to deal with spatial patterns by themselves are confined to the image, neglecting the realities which it reflects.

In the first section of this part, *Location and Spatial Structure*, we try to convey a sense of the principles governing the spatial organization of a society. Rutledge Vining provides a powerful intuitive glimpse of the way in which the peaks and valleys of a socioeconomic landscape organize themselves (chapter 1). The mechanics of his approach are not as important as the sense he conveys of chance and causation in a cumulative process. In contrast to this macrogeographic view, William Alonso provides a microview of the location decision of the firm. In his chapter "Location Theory," he summarizes the basic ideas of the classical theory of economic location (chapter 2). Although the limitations of this theory are well recognized today, it still constitutes an important perspective on the organization of spatial phenomena. In a second contribution (chapter 3), Alonso extends classical location theory to allow a broader look at the many other conditions of locational choice. These operate most strongly in developing countries, but they play a role even in highly developed economies. August Lösch, who, together with Walter Christaller, may be considered the founder of modern spatial theory, presents a glimpse of one of the fundamental contributions to the field (chapter 4). His concept of a nested hierarchy provides an architecture of great intellectual elegance which, together with later extensions, underpins central place studies, particularly with regard to the distribution of population-serving activities. In a more descriptive vein, Brian Berry's chapter, "The Geography of the United States in the Year 2000," is a piece of synthetic future history based on much present knowledge of structural theory and trends (chapter 5). Although any exercise in social forecasting has, as we know from experience, a low intrinsic probability, Berry presents an intriguing scenario. His approach echoes Rutledge Vining's.

The second section, *Spatial Systems in Economic Development*, stresses the dynamics of movement and economic growth in the national spatial



system. Introducing this section is Albert Hirschman's now classic essay, "Interregional and International Transmission of Economic Growth" (chapter 6). On the basis of an experienced intuition, he suggests a cycle of increasing concentration and inequality in regional distribution during the earlier stages of economic development, followed by a progressive spread of development in later stages. Empirical confirmation of this thesis comes from Jeffrey Williamson (chapter 7), who offers extensive historical and cross-sectional documentation of the pattern. Complementing these long-run developmental views, Wilbur Thompson presents for the United States a concept of continuous creation of economic activities at major centers and a continuous diffusion of these activities to the periphery (chapter 8). With Peter Morrison's contribution (chapter 9), we shift attention from the location of economic activity to changes in the location of people, or migration. His chapter brings together some of the most recent evidence on the dynamics of population movement in the United States, much of which challenges traditional concepts. In the following selection, Peter Gould, a geographer, delineates the changes in spatial organization of the Tanzanian economy over a period of four decades (chapter 10), showing the orderly evolution of a spatial system in which the past exerts a powerful influence upon the future. Finally, John Friedmann discusses three concepts of power and their effects on spatial structure in developing countries: political or decision-making power, environmental control attendant upon innovations, and economic power in a multiethnic social context (chapter 11).

In the third section, Growth in Subnational Regions, we present some now classic views of the basis of economic growth in particular regions. Harvey Perloff and Lowdon Wingo review the American experience of regional growth in relation to resource endowments and economic specialization (chapter 12). They conclude that the role of resources can be understood only in a broad context of spatial interaction, changing technology, and shifting patterns of demand. Their use of a heartland-hinterland concept anticipates the theory of spatial dualism that underlies much current theorizing in the field. In the following two chapters, Douglass North and Charles Tiebout argue out the role of exports in regional economic growth (chapters 13 and 14). These pieces are among the oldest in this collection. They have been retained primarily because it is interesting to see ideas being hammered out. The integration of regional export base theories with general Keynesian theory has long been accomplished, and the interdependence of the regional multiplier and regional exports is clearly understood, so that strategies of import substitution are often suggested as alternatives to the expansion of exports for regional economic growth. Nonetheless, export-multiplier approaches are still rather mechanistic, and the question of what determines dynamic expansion and innovation is very much alive today in more developmental terms. In some cases, it is argued, multiplier effects do not take place because of the lack of certain social and economic forms of organization, so that regional exports can drain a regional economy as readily as they can stimulate

its further growth. In the final selection (chapter 15), William Nicholls examines the bases of diverging rural economic well-being in small regions within the American South and finds that local industrialization, far from conflicting with rural prosperity, affords complementary and reinforcing relationships within a well-defined spatial context. The reader should know, however, that subsequent studies in this vein have revealed a variety of patterns according to local circumstances.



---

## **Location and Spatial Structure**



# 1 An Outline of a Stochastic Model for the Study of the Spatial Structure and Development of a Human Population System

Rutledge Vining

IN PART I, A description is given of the concrete object whose structure and development are to be studied. The figure that I use for this purpose of conceptualizing the phenomenon has been submitted previously in published papers.<sup>1</sup> It is included here as an aid for a reader's visualizing the observable aspects and features of pattern and form for the analysis of which a method is being proposed.

I have tried in this part to present a mode of description that one conceivably could apply actually and literally by duly marking and arranging objects of some sort representing the elements of an existing population, the marking to be done on the basis of information that one may imagine being provided by a feasible census. In outlining what one may do with census-type data to obtain a physical representation of a *population system*, I have included also a number of statements indicating conjectures as to what he would see before him as the result of his following the instructions. These conjectures are more or less general impressions of mine based upon a fair familiarity with the descriptive literature, and I refer any curious reader to the excellent and comprehensive bibliography and commentary, *Central Place Studies, A Bibliography of Theory Applications*, prepared by Brian Berry and Allen Pred and published in 1961 by the Regional Science Research Institute.

Having indicated a mode of describing an existing population system, I try in Part II to illustrate the peculiar kind of inferences that men have drawn from observations made upon certain aspects of the discrepancies, divergencies, and general features of variation characteristic of a development of this kind of system. These inferences are of the sort that lead one to assert or imply that differences between observed values of defined variables are not what they *should* or might better be. One observes a particular pattern of variation and is induced thereby to declare as faulty the performance of the currently operating economic system, this being something that is subject to supervision and alteration by legislators. And such a deduction or declaration implies some

---

Work upon this paper was supported by a grant from Resources for the Future, Inc., and I wish to express appreciation for this assistance.

<sup>1</sup> "Delimitation of Economic Areas: Statistical Conceptions in the Study of the Spatial Structure of an Economic System," *Journal of the American Statistical Association*, March, 1953; "A Description of Certain Spatial Aspects of an Economic System," *Economic Development and Cultural Change*, January, 1955; "On Specifying Frequency Distributions Descriptive of the Outcome of the Working of an Economic System," unpublished, 1960.

Reprinted from *Papers of The Regional Science Association*, vol. 13 (1964), pp. 15-40, with permission of the author and The Regional Science Association.

version of an "expectation"; for to say in effect that the observed values of variables, or the differences between and among observed values, are too large or too small, or in some way indicative of a faultiness or deficiency about how something is working, is to imply that one can recognize values of variables (or differences, divergencies, and discrepancies) that are not too great or too small: that is, that there are "expected" values of the variables in question or an "expected" pattern of variation against which one may gauge the variation and values observed.

This concept of *expectations* being probabilistic in nature, the main section, Part III, follows in which I attempt to give the bare outlines of a probabilistic model, again in such simple terms, that one might conceivably perform the experiment by hand. Hand-performance would of course not be practicable, but I am given to understand that computers can be programmed for the performance of such an experiment. If so, then by Monte Carlo procedures one might take a small step toward defining a concept of "expectations" which would seem to be implied in statements to the effect that an observed value of this or that variable (say, average family income, or inequality of distribution of income, or rate of growth, or degree of industrial development, or degree of "economic balance," or degree of "economic stability," or whatnot) is too low, or too high, or not enough, or excessive, unexpected, and such as to induce suspicions of something's being in an unacceptable state.

#### PART I

As indicated above, this part is presented as an aid for a reader's visualizing the concrete object whose structure and development are to be analyzed.

Imagine a large area map of this country, and suppose that upon this map a poker chip has been placed at the location of the domicile of each of the millions of family and single units of the population of this country. The millions of chips so placed would form something that may be called a density configuration. The pattern represented by the variation of densities could be made more distinct perhaps by arbitrarily dividing the entire area into many small sub-areas—say 10 or so miles square—and putting the chips contained in any one sub-area one upon the other to form a column.

The spatial distribution of columns would represent a configuration of densities. Just as an ordinary frequency histogram shows the densities over the designated intervals of the range of a variable, the heights of the stacks of chips would show the densities of individuals for the designated sub-areas in two-dimensional space. That is, for any arbitrarily determined origin and coordinate axes, the columns of chips would describe a two-dimensional frequency histogram, each sub-area being located as a neighborhood about a point in space and having associated with it some frequency.

In passing, we may remark generally upon what one would see upon the map. Each area or group of areas containing a city would show up as a density peak. One of these would be seen to tower over all the others. Then there would be several scattered over the nation which, although half or less as high as the tallest, yet tower over all the immediately surrounding peaks. Scattered

about these would be lesser peaks, and about each of these lesser peaks there would be peaks of still lesser heights. Most of the sub-areas would contain no chips or only a very few and would represent areas containing no towns. Of the remainder, a relatively large number would each contain small numbers of chips, each such area representing the location of a village or small town. A small number of areas would each contain a relatively large number of chips, these sub-areas representing the locations of sizeable cities.

The primary frequency histogram to which the above remarks refer represents a classification of family and individual units with respect to the coordinates of their respective locations. As a start toward specifying and describing whatever pattern or shape that may be discerned in the spatial orientation of family units, let us suppose that a derived, single-dimensioned frequency distribution were constructed, this one depicting the variation, not of the number of families with respect to their locations, but of the number of sub-areas with respect to their heights of columns (i.e., to their respective densities). Suppose that this second frequency distribution were formed by counting the number of sub-areas containing exactly  $X$  chips for each possible value of  $X$ .

To do this manually, one might conceivably follow some procedure whereby a tallysheet would be used. Suppose the tallying were done by one's randomly drawing an initial sub-area and then taking the rest in some kind of order of proximity. Beginning with the randomly drawn place, one would proceed with those sub-areas immediately around this initial area, then with those immediately about these, and so on,—so that as the number of sub-areas tallied is increased a larger and larger aggregate and connected area would be included in the sub-areas tallied.

Let us assume that when one follows this latter procedure he finds not only that a relatively simple mathematical formula describes the distribution for the total area but also that the observed distribution on the tally sheet, whatever be the initial sub-area that is randomly drawn, more or less rapidly approaches a stable form with an increase of the number of sub-areas tallied. That is, suppose that for each of two halves of the total area the distributions are approximately similar; that for each of four quarters of the total area the distributions are likewise similar; that for each eighth, et cetera; or, generally, that any part of the total area containing a relatively large number of the sub-areas would show a size distribution of communities similar in form with that of any other comparable part.

Now if this were so, the regularity represented by the stable form of this distribution would constitute an element of a pattern or shape that one would observe in the spatial orientation of the columns (population concentrations) upon the map.

In the above paragraphs we have spoken of only one aspect of this object which we are calling the human population system: the locations at an instant of time of the component individuals and family units. In a description of this aspect, the chips representing the individuals are undifferentiated. But there are untold ways of classifying and differentiating the individuals, and let us now consider some of the more common classifications.



The spacing of individuals and their clustering at central places are a spacing and clustering of complementing activities. Each individual represented receives a money income by contributing to the production of certain goods and services. Suppose first that the chips are differentiated with respect to the type of occupation or service constituting the source of income.

Many of the chips represent population units whose incomes are derived from the gathering of natural resources or the using of nontransportable natural resources: farmers, miners, loggers, commercial fishers, et cetera. These chips are thus dispersed over the entire area in conformance with the dispersion of the various natural resources that are being exploited. The density peaks represent clusterings of many kinds of activities pertaining to the assembling and processing of the products of the dispersed individuals and to the providing of services to neighboring dispersed individuals as well as to assemblers and processors. Now inasmuch as the uniformities and regularities that students purport to see in a spatial arrangement of a human population system are largely the same whatever the time period or whatever the continent over which the population is spread, there is at least a suggestion that for sufficiently large regions there is some kind of uniformity among areas with respect to the spatial dispersion of natural resources. I shall attempt in a subsequent section, as a part of the outline of the stochastic model, to define a special sense of regional uniformity of natural resource endowment. The definition is probabilistic in nature, so that deviations and variations are not inconsistent with the assumption of this special kind of uniformity among large regions with respect to natural resource endowments and with respect to the spatial dispersion of these resources. In this section we shall simply suppose that there is a uniformity. Individuals who exploit and mine these resources are of course similarly dispersed. And the chips representing these dispersed individuals could be differentiated in some way or another.

These dispersed individuals are serviced and supplied from central places, these latter being represented by the columns on our map. Again with reference to what is assumed as general knowledge, one may say that the individuals who compose the populations of the smallest central places typically produce a limited number of kinds of consumer goods and services, each being characterized by a small market range,<sup>2</sup> a high per capita consumption, a small producing organization, and requiring only generally prevalent skills for its production. In addition, there are individuals in these small places who are engaged in the assembling and shipping of the products of the dispersed individuals. Let us suppose that the chips representing these individuals in the smaller towns can be differentiated in some way to indicate the type and grade of occupation constituting the source of income.

<sup>2</sup> In brief, the notion of "market range" is this: each unit of value-added of a particular kind of service or product is consumed by an individual or agency at some point in space. The headquarters for its production is located at some other point in space. For each unit, then, there is a corresponding distance between these two points; and "market range" refers to some kind of average of these distances for all the units of that good or service.

Continuing with what I shall assume to be common knowledge, several of these small places would be found scattered about a somewhat larger place. The individuals in this larger place produce, in addition to the kinds of goods and services produced in the smaller places, goods with greater market ranges. Also, rudimentary processing operations are more likely to be found, and perhaps there are individuals who supervise the activities of individuals located in several of the smaller places. Thus, there would be chips in the column representing this next-sized place that are not differentiated from chips in the smaller column, but, in addition, there are chips with differentiations not found in the smaller column.

The same may be said about the next larger place, and the next, and the next. The columns for the large cities would show many differentiations among the chips, and the number of these differentiations declines with the size of the city. Whatever the size of the place, there would be occupations in it similar, with respect to grade and type of capacity required, to the occupations found in any of the smaller places ranged about it. But the larger the place, the more the kinds of products and services produced, the greater the proportion of individuals engaged in administering the operations of individuals located in smaller places, the higher the quality of the professional services available within it, the greater the number and the more distant the places from which products are assembled, the greater the number and the more distant the places to which products are sent.

Suppose now that the chips were differentiated in some way—perhaps by the sizes of their diameters—so as to indicate the amounts of the money incomes of the corresponding family and individual units. We know, of course, the pattern that would be evident. As has been said, cities of a given size contain individuals who perform the same kinds of services and exercise the same kinds of talents and capacities as those performed and exercised by individuals who compose cities and towns of smaller size. But, in addition, the larger city typically contains individuals performing administrative and professional services that are not found in the smaller communities and that are of a relatively high value; and also, even in performing services ostensibly similar to those available in small places (e.g., medical), there are individuals in the larger centers who exercise capacities and talents of a quality that is higher, and thus more highly remunerated, than those typically found in the smaller. Moreover, in the larger city the producing organizations are large and more mechanized so that the amount of capital employed per individual is larger than in small cities and towns; and the domiciles of the persons whose incomes are supplemented by the yield of the properties in these accumulations of capital are predominantly in these larger cities. And then there is this important consideration: the high density of population that favors the growth of relatively large sizes of productive organizations also favors relatively high degrees of division of labor and specialization upon goods and services consumed locally. The larger the city the greater the extent of this division of labor and specialization; and, accordingly, less of the goods and services consumed by individuals in the dispersed areas and small towns is exchanged and thus evaluated

in terms of money than is the case in the larger city. In this sense, we may say that the degree to which the activities and products of individuals are "monetized" is least in the dispersed areas and increases with the density of individuals or size of cities. The extent to which the division of labor has been carried and to which consumer goods and services are exchanged for money being greater in the large city than in the small towns and dispersed areas ranged about it, a given level of consumption of real goods and services corresponds to a somewhat higher level of money income in the large city than in the towns and outlying areas.

All of which would suggest—and the available data are clearly in accord—that the larger the city the higher the average income of the inhabitants. Thus, were we to start with one of our taller columns on the map and consider the sub-areas surrounding it—say by taking groups of sub-areas in successive concentric belts—we should find, independently of the location of the tall column in question, that average family income declines as distance from the center is increased until a nearby tall column begins to be approached. Average family income declines markedly as one moves out from any large city.

This differentiation of the chips—however it may be effected, by color or otherwise—with respect to the types of occupations and qualities of capacities and money incomes of the individuals represented, would show an additional aspect of the pattern and form evident in the population system. This aspect would be observed in the characteristically different compositions of types of activities, qualities of capacities, and variety of talents among the columns of varying heights (the various sizes of cities and towns). Suppose that columns of similar heights, when the component chips were differentiated as suggested (say different colors and shadings for the different types of activities and occupations, with sizes of diameters made proportionate to the amounts of family income), did in fact show similar color compositions and shadings and approximately the same average diameters. Then, there would be this additional aspect of form and pattern discernible by eye—the tall columns distributed over the larger areas all having about the same colorations and shadings and about the same average diameter of component chips, and there being about the same kind of shading-off of color and diameter size as one's eye moves out from any particular tall column.

As in the case of the aspect of pattern previously mentioned, perhaps this visual impression could be quantified in terms of a few relatively simple formulae. For example, consider again the respective average family incomes calculated for one of the density peaks and for each successive concentric belt of sub-areas around this tall column. Suppose when these averages were plotted against distance from center on the horizontal axis that a relationship is shown that may be approximated by some relatively simple mathematical formula, and that this relationship looks to be about the same regardless of what particular density peak is selected for the plotting. And suppose further that when other calculations are made for each column—say the proportion of chips representing individuals whose occupations are in manufacturing industries, or the per capita outlay upon goods and services publically provided and locally con-

sumed, et cetera—one sees in the plottings regularity and stability of form that may be approximately described by relatively simple mathematical formulae.

If this were so, as there is some ground for thinking, the regularity represented by the similarity, among the various tall columns, of these stable forms of relations would constitute additional elements of pattern and shape that one would observe in the spatial orientation of the population concentrations.

Consider now one further aspect of the conceivable features of form and pattern in the population system. Suppose a second map were taken, with the same sub-areas as those lined off on the first, and let us imagine placing upon this map chips to represent, not individuals and family heads, but units of value-added by production. On the first map, a tall column represents a relatively large central place, a clustering of domiciles of individuals. During a given period these individuals will have produced some number of units of value. In the course of time, these units of value are consumed in some roughly definable sense. Suppose that data were available indicating the domicile of the consumer of each of these units of value. And let us imagine that for each unit of value produced by individuals living in this particular central place a chip is placed in the sub-area containing the domicile of the consumer of this unit.

As in the case of the first map, there of course would be a density configuration represented by the columns of chips distributed over this second map. And we may make tentative statements about what this configuration would look like. The tallest column would be in the sub-area in which the central place itself is located, a very large part of the value output within any central place typically being consumed close at hand. And as one moves out from this center, the heights of the columns of chips on this second map, just as those of the columns on the first, would rapidly decline. If in some way the chips were made distinguishable by center of origin, the destinations of the value outputs of the inhabitants of more than one central place could be represented on the same map. And for centers that are far apart, there would be little intermixing of these differentiated chips, for only a small part of the value produced within one central place is consumed within central places that are far away. But with central places that are relatively close together, there would be an intermixing of the chips and an overlay of the respective density configurations.

Just as one might locate upon a map of this sort the destinations of value output that originate in a given center, so might the originations of the units of value consumed within the center be located on still another map. And if such were done, the distribution of chips about a given center would be closely similar; the spatial arrangement of the columns of varying heights would give an impression of about the same pattern for the two arrangements: the destinations of units of value originating in a given center, and the originations of units of value having a final destination in that center. A major part of the value consumed within a center is also produced within or closely around the center. And as one moves out from the center, the density of these chips

representing the origination places of the units of value consumed would show a rapid decline.

One may note that whereas the map and density configurations discussed in the first sections of this part represent phenomena that have their existence at a point in time—the locations of existent, classifiable objects represented by differentiated chips placed in columns—the second type of map and density configuration represents cumulations of flows. Actual things will have actually moved from place to place, and that which would be shown on the map refers to aggregations over some definite period of time. The distinction is the familiar one between stocks and flows.

There are many kinds of such flows to which attention may be drawn, e.g., money payments and receipts. During any ordinary period of time an individual makes many payments, and each such payment for all the individuals represented in a given column on our first type of map has a destination in the location of the receiver of the payment. Suppose that the locations of an individual or agency and of this individual's or agency's deposit bank were the same and that all payments by individuals and nonbanking private and public agencies were made by check—with bank reserves being adjusted as they now are through transfers in the gold settlement fund. Then we should have a straightforward and unambiguous accounting of all money payments: the origination and destination of each dollar paid and received. Accordingly, maps of our second type could be formed for these money payments and receipts, and these maps would be derived from information on quasi-physical monetary flow phenomena—indeed, strictly physical in conception, the deposit accounting being an efficient substitute for specie flows along identifiable lines connecting the various central places.

This that would be observable as a density configuration on our second type of map may be quantified, as in the case of the first type, in terms of various forms of derivative frequency distributions. For example, corresponding to each unit of value-added or money payment there is a distance between the point of origination and the point of destination. Let this distance be measured for each value-unit, and suppose that for all value-units originating within a given sub-area, or group of contiguous sub-areas, a tally were made of the number of value-units whose distances-traveled are less than  $X$  for all possible values of  $X$ . If this distribution were such as may be approximated by some relatively simple formula—say a logarithmic normal distribution of some linear function of  $X$ —and if the parameters of this distribution were approximately the same for similarly sized central places and relatively stable over time, then in the regularity of this distribution form and stability of parameter values would be manifested another element of pattern and shape characterizing the population system.

The stability of form and pattern which we may imagine (with some support from empirical studies) as observable on our first type of map, refers to the *population system* per se as a distinct entity and not at all to the component elements of the population. At a point in time there is a density configuration and over relatively long periods of time certain features of pattern—specified

in terms of forms and parameters of frequency distributions—are supposed to remain approximately the same. But the individual members of the population are in a constant state of motion, and the individual columns representing central places are continually varying in height and composition. Considering the happenings over any period of time, we know that some individuals die and others are born. In regard to any central place and during any period of time, there are families moving out and other families moving in. In all places, there is a continuing birth-and-death, adding-to, and dropping-out process at work. But throughout all of this internal and individual flux, the population system as an entity is maintaining a kind of stability.

One may readily imagine a sequence of maps of our first type, say one for each year over a long sequence of years, that would demonstrate the continual motion of individual elements of a population system which itself is maintaining the kind of stability referred to. A column of chips representing a central place as it would stand today would consist of chips some of which would have been found in other columns on the map depicting the situation of a year ago. Suppose that each column of chips were assigned its own number, and let each of the component chips of a given column bear the number of the column containing the chip one or three or five years ago.

Students of migration would know approximately, on the basis of census data and sample studies, what to expect regarding the composition of these numbers within the given column. A relatively large proportion of the chips would bear the number of the column in question; a smaller proportion would bear the numbers of columns in the concentric belt of sub-areas contiguous to the sub-area of this column; and for successive concentric belts this proportion would decline rapidly. The pattern, regardless of the particular column selected for study, would be much the same for similarly sized central places and would appear in general features much the same as that shown on our second-type maps.

This time-sequence of maps now being considered, and the second-type maps considered above, would be derived from information upon physical or quasi-physical flows. Actual things are continually coming-into and passing-out of existence; of course, the latter rate being typically less than the former, the aggregate number of elements contained in the population in question is increasing with time. And these actual things are moving about, passing along lines connecting the various central places—as one literally sees as he looks down from an airplane. This growth in numbers on the one hand and continual motion on the other represent the life and development of the entity which we are calling the population system. The development is manifested in the variation over time of the respective heights of the columns representing the sizes of the central places and also in the compositions of qualities of talents and types of occupations and activities of the component chips of the various columns. This gradual alteration of heights and compositions is the form that the growth and development of the system assumes, and among the different parts of the system there is variation of rates of growth and development which may be depicted, as in the cases of other variation phenomena referred to above,

by forming frequency distributions.

From the data and information that would be contained in any successive pair of maps in the sequence that we now contemplate, a rate of change of size (height of column) could be computed for each central place. The computed rates for all the central places could then be formed into a frequency distribution for each interval of time, and, for any sequence of time intervals, there would accordingly be a sequence of these frequency distributions of rates of growth. To the extent that these distributions exhibit regularity of form and stability of parameter, as there are grounds for supposing, there would be manifested here an additional aspect of pattern characterizing the *system*, this one pertaining to its development over time.

In sum, transcending the variation among rates of growth and development and the vast complex of flux on the part of the individual elements of the population, there are stable features of form and pattern characteristic of the system *per se*. The individuals are not "at rest," nor are they ever "tending to a state of rest." They are in a continuing state of motion and change. But at any point in time they are distributed over the total area, and counts may be made for each of any set of arbitrary small areas into which the total area may be partitioned. One may think of this distribution as tending to maintain a regular and stable form. This is the kind of situation described by the idea of *statistical equilibrium*, and we shall return to this concept in Part III below.

## PART II

The concreteness of detail and simplicity of method of description for which I have striven in the preceding part tend, I think, to serve the good purpose of clarifying the nature of that which can possibly be observed as the outcome of the working of an economic, social, and political system. I shall try to explain how, and to this end I wish now to place main emphasis upon a distinction between (1) the object familiarly referred to as an economic system which men regard as immediately alterable by deliberate decision and direct choice exercised jointly by functionaries called legislators, and (2) the performance characteristics of a particular instance or design of this object. There is a thing that performs and there is the performance of that thing, and I wish to keep distinct these two phenomena.

In regard to the first of these two things, I shall ask a reader to recall some occasion when actual persons were discussing and deliberating upon the question of whether or not the economic system, or some component thereof such as the monetary system, is working or performing as well as it may be made to perform. Typical discussions of this type are recorded in parliamentary and congressional hearings upon what is spoken of as economic legislation. For example, the hearings which led to the design and subsequent adoption of the Federal Reserve Act (or any subsequent hearings on proposed amendments of this statute) constitute a sort of dialogue on laws in which responsible men are discussing among themselves the performance of a component of the economic system and considering ways of altering this object for the improvement of its working. So, too, is this the case for hearings on the farm problem, or the

monopoly problem, or the labor problem, or the problem of depressed areas, or any of the many others that may come to mind. The discussants in such recorded dialogues consist (1) of experts and specialists of various sorts acting as advisers and expositors of fact and explanation and (2) of persons who will have invited the specialists to discuss with them various questions that have been raised in connection with a choice that these persons are obliged to take a part in making. The choice or decision that is to be subsequently exercised is between the economic system as it currently stands and the system as it would be altered by someone's proposed design of a modification. The persons who directly and immediately exercise the choice are legislators. The proposed modification invariably takes the form of a draft of a statute whose adoption would amend the present system of statutory law and administrative rules.

Now if it is sensible to speak of these persons as deciding whether or not to adopt a proposed modification of the *economic system*, and if, in fact, what they are literally deciding upon and choosing is a system of statutes and administrative rules of action, and inasmuch as a statute is a set of constraining and prescriptive rules of action establishing personal roles to be played and imposing conditions upon the exercise of private choice, then it is sensibly in keeping with observed usage for one to understand the term *economic system* (to the extent that its reference is restricted to an entity deliberately alterable by a people's acting through legislators) as referring to the system of legislatable rules of action which set conditions within which individuals exercise their private choices.

So much, then, for the thing that does the performing. Call it what one will, it is a system of constraining and prescriptive rules, analogous to "rules of the game," that real persons actually choose when they are conscious of choosing what they speak of as an alteration or modification of the economic system. To describe or specify an alteration of the system is to set forth an amended set of constraining and prescriptive rules of action, just as to describe a deliberately chosen alteration of a game is to set forth amendments of the rules of the game. Note especially that we did not describe an *economic system* in the preceding part, where we described a quasi-organic entity which we named population system. An *economic system*, in the sense that seems to conform with the behavior of persons charged with responsibility for the way it works, is a set of rules implicitly understood and accepted by individuals as commands constraining and conditioning their respective private choices and decisions.

The system in question being a system of rules, consider now the matter of describing the *performance* of the system—how it has been working, or how it may be made to work if only certain alterations are made. If one again recalls an occasion of an actual discussion of the sort we now contemplate—among men responsible for reviewing the performance of the economic system and for altering this object to the extent that its performance is expected to be improved thereby—he may see what is done by a person trying to communicate to others what he observes as the essential aspects of the performance of the system. Much of these discussions have to do with presentations and



explanations of tables of numbers, charts, and diagrams upon which lines and bars are drawn. The tables and charts are used to illustrate the behavior of selected variable quantities. The persons presenting these data and exhibits think and speak of themselves as demonstrating to the others how the system has been working—not only how it happened to have worked during a specific period of time, but the working properties or performance characteristics of the system per se as it currently stands.

Here, then, are the two things: the system of constraining and prescriptive rules of action such as may be amended by statutory law and the performance characteristics of a system, as it exists at some point in time or as it would be altered by the adoption of some proposed amending statute. Bear in mind that it is the system of rules that is subject to direct choice and alteration rather than the expected values of the variables in terms of which the performance of a system is described.

We are all familiar with the variables typically represented in the tables and charts submitted in legislative discussions: prices of various commodities, averages of prices computed for various classes of commodities, outputs of commodities and services, and averages computed for various classes of these articles; aggregate income of families and individuals, averages of incomes computed for various classes of families, and distributions of family incomes variously classified; total numbers of persons employed and unemployed, and proportions of persons employed and unemployed computed for various classes of individuals; proportions of persons employed in various occupations; amounts of money in the possession of various classes of individuals and firms; rates of growth computed for a variety of variables and for various area classifications. These are typical and there are many others that one need only consult a record of such discussions to see.

Let us be reminded now of a peculiar feature of these discussions, or rather, of a specific detail of the natural behavior of the discussants—natural in the sense that this type of discussion would seem to imply as unavoidable this feature or detail.

The discussants are observing and describing the performance or working properties of the economic system. In doing so they point to and talk about some particular pattern of variation: certain variables will have done this and others that, or certain variables will follow this or that pattern if some proposed alteration of the system were adopted. And their talk and explanation extends to a typical kind of inductive inference. The observed pattern of variation will have induced some of the discussants to a belief that the present system is not performing so well as an altered form of this system would perform: that is to say, that *faultiness* is indicated in what can be observed. The inference is reached somehow that there are changes that can be feasibly made for the improvement of the performance of the economic system. And this poses questions: How are the variables selected for the observation and description of *performance*? How can a characteristic pattern of variation, observable in some arrangement of the values of these variables, be derived as an "expectation" of the outcome of legislators' choosing some specific modifi-

cation of the *system*? How is a person induced by what he observes, in the bare form of a pattern of variation, to a belief that a system is performing less well than it might be made to perform?

It is the last question that is the most intriguing to me: How is it that one gets an impression of *faulty* performance when he looks at a particular pattern of variation? How can "too much" or "too little" be discerned in what one may see? These qualities are of course never directly observable. But one may note many references to them as though their recognition scarcely presents a problem.

I have in mind now a case of an interchange between a congressman and a prominent economist that illustrates what is yet left to be done. In this instance, we may see a legislator acting for all the world like an operator of a quality control system, observing plotted values of variables upon a chart, and inferring from his observations that the thing for which he bears responsibility is not working properly. The chart shows a number of lines indicating what has been happening to each of several variables. The legislator points to two lines in particular that are "going farther and farther apart." His remarks suggest that he is taking for granted that the pattern of variation exhibited on the chart indicates faulty performance of the present system of statutes and administrative rules. And his demonstration and comment seem hardly more than a brief prefacing for his embarking upon the main task, viz., the designing of repairs and modifications for the fixing of this faulty performance.

But to the astonishment of the legislator, the economist declares, with reference to the variation pattern on the chart, "That is not an abnormal thing that one sees up there."<sup>3</sup> That is, the specialist sees the variation shown upon the chart as being what one would expect of any admissible system and hence as not indicating faulty performance. The congressman is nonplussed and replies, "You mean this is not abnormal. You surely are not saying that one line should run one way and the other line the other." Whereupon the specialist tries to explain the kind of variation, among the variables shown on the chart, that reasonably might be expected when an admissible system of enacted constraints and prescriptions is working properly, normally, or characteristically.

This incident appears to me to be an apt illustration of the kind of situation that has given rise to the development of the conventional form of theoretical economics. From this point of view, the classical version of theoretical economics is in essence a theory of economic organization, and its systematic development over the years has been the result of men trying to explain such judgments and statements as those of the specialist and trying to respond to such questions as those of the congressman. The objective in the minds of those primarily responsible for the theory's formulation has not been, in the main, the instructing of individual economizers in the exercising of their private choices of modes of production, rates of output, rates of consumption, et cetera, but rather the analyzing and explaining of the effects of observed or

<sup>3</sup> The congressman is Representative Patman, the economist is T. W. Schultz, and the occasion is reported in the *Hearings before the Joint Committee on the Economic Report*, January, 1955, pp. 624-43.

proposed changes in the conditions under which individual families and firms respectively make their private choices. The theory is applied to describe these effects, and the description is given in the form of statements indicating what would tend to happen, as a result of a specified change of conditions, to each of a set of variable quantities. It is presumably an application of this theoretical model that leads the specialist to declare in this instance that "There is nothing wrong in what one sees up there." For one may infer, by this method, allowing for the changes in technology and consumer preferences that are understood to have occurred, that "What one sees up there" would be seen whatever system of statutory and administrative rules of action is in force, provided only that individuals are left some choice regarding what and how to produce and regarding the rates at which they exchange whatever they are allowed to trade. Inasmuch as "one line going one way and the other going the other" is what one may normally expect under the prevailing circumstances of a system of constraints that is working as it should work, the appearance of this phenomenon is not in itself indicative of faulty performance.

But note wherein this argument is unsatisfactory from the congressman's point of view; for his doubts about the goodness of the performance of the system of statutes as presently administered have been induced, it appears, not by his observing merely one thing being less than another or one rate being smaller than another, i.e., not by his simply observing a lower end of a range of variation, but by his observing what has seemed to him to be an exceptional shape or form and an abnormally large amount or size of variation, divergency, and discrepancy. The form has seemed unusual and the size or quantity of variation has seemed "too large" to him. And the above argument does not get at this source of doubt, there being no consideration of the "expected" form of the variation nor of a criterion for judging the largeness of the observed amount of variation.

Moreover, it is substantially this same kind of virtual comparison that leads the specialist himself to believe that the system is working badly. No less than that of the congressman, the behavior of the specialist, when he turns to a consideration of his own observations, is also suggestive of that of an operator of a quality-control procedure. He, too, finds occasion to describe what he has observed and declared to be unusual and "excessive" variation but pertaining to another selection of variables. While accepting the variation described by the legislator as within the bounds of "expectation," he points to observations upon another set of variables and infers therefrom that the system as it stands calls for modification. The discrepancy between small and large family incomes is "too large"; the concentration of very low family incomes within particular geographic areas is "too great"; the persistence over time of this "excessive" area concentration is "too long." "Size of discrepancy," "degree of area concentration," and "temporal persistence of a discrepancy" are all so large as to indicate faulty performance of something subject to choice by the joint action of legislators. It is, of course, not maintained that one may infer faulty performance from an observation of variation among the respective rates of change of average incomes of various classes of families. As in the case of the quality-

control operator, one would expect some variation among the observed values of the variables when the system is working as it should work. But the observed variation is declared to be "excessive": low-income families are observed to be geographically "too concentrated"; the rates of movement of families from low-income areas to high-income areas and from low-productive activities to high-productive activities are "too small"; the divergency of growth rates of per capita output and income for the various areas over which the population is spread is "too great."

One may readily sympathize with such judgments; the discrepancies, divergencies, persistences that one observes do seem large in some sense—somehow larger than they should or ought to be. But as I reflect upon what may be going on in my or someone's mind when this sensation is experienced, I am not at all clear about the inductive inference that is being reached. This, I submit, is the intriguing question that is raised for specialists to work upon: how is it that one gets an impression of *faulty* performance when he looks at a particular pattern of variation? How does one recognize "excessiveness" in an observed pattern of variation? This would appear to be a neglected class of analytical problems. To discern "excessiveness" in the variation that is observed is to see an "unexpected" shape or form of a frequency distribution. And this implies an "expected" form or shape with which the observed form is compared. In the case of the simple inferences prescribed by the rules followed by an operator of a quality-control procedure, or by anyone testing a null hypothesis or judging the significance of a difference between observed values of random variables, a standard of comparison is explicitly specified. The corresponding specification for our case is missing.

It is at this point that it seems to me the kind of description presented in Part I above becomes relevant. In that part we tried to present something that could be literally seen in visible shape and shadings—a pattern formed of variously sized and spatially oriented stacks of differentiated objects. Conceivably, one could actually construct the thing for each of a long succession of points in time and thereby see real evolution and development of a pattern of form and color. Nothing was said about anything that could not be represented by a placement of a differentiated chip.

Now, if one should search back through Part I, he would find a considerable number of the variable quantities to which students of economic phenomena refer when they describe the working or performance of the economic system. For example, national income and product, industrial and other classifications of output, total quantity of money, total volume of money payments, per capita incomes computed for various area classifications, and the like would all be available as aggregates of data indicated by the markings on the individual chips; similarly for occupational structures and changes thereof and for the various familiar measures of rates of growth, rates of urbanization, rates of industrialization, et cetera. Discussions recorded in economic journals and in hearings before parliamentary and congressional committees largely consist of statements about such variable quantities: what has happened, or is about to happen, or will or would happen to them unless or if certain legislative or ad-

ministrative decisions or choices are made. And note that the language used in the statements is such as to imply a judgment on or evaluation of the amount of variation being described by the statements. One need only recall familiar comparisons by states of per capita incomes, value-added by manufactures per man-hour, proportions of employed man-hours occupied in manufacturing industries, expenditures per capita on government services, rates of change of population, income, industrial output, levels of aggregate income, employment, and unemployment and quarter-to-quarter changes thereof. We can all call to mind instances of the very fact of observed differences, discrepancies, or divergencies being presented as indicative of something's being wrong with the working of the economic system.

And yet a simple placing of chips upon an area in accordance with such instructions as those outlined in Part I leads one to a contemplation of a kind of pattern of variation that somehow seems "natural" or normally to be expected. We really would not expect, even though we are unable to be explicit regarding precisely why, all communities to be the same size. We feel comfortable with variation in this respect, and what's more we expect small communities to be scattered about the large ones in a way suggestive of a roughly hierarchical arrangement. Moreover, we don't expect the same distribution of talents and qualities in the small as in the large places, nor the same per capita incomes, output of manufactures, expenditures on governmental services, et cetera, in the small as in the large. We intuitively expect a quite marked systematic variation among the occupational structures of the many communities. This intuitively expected systematic variation would be shown on such a map as we considered in Part I.

The map did not include boundaries of states or counties. The sub-areas were arbitrary, and there would be ten or twenty thousand or so of them for a nation so large as ours. But now suppose we were to construct a kind of wire grid or net so that when tossed over our map the entire area would be partitioned into about fifty regions, each including a relatively large number of the sub-areas but all not necessarily being the same size. These fifty regions would be no less arbitrary than are the sub-areas and would correspond to something like our states. And, if we may rationalize as plausibly to be expected, the variation indicated on the map as it stood, then the variation among aggregates and averages computed for the fifty larger areas similarly would be plausibly expected.

Now if we can be led to take this step, I submit that we will have gone at least a small way toward formulating a conception of "expected variation," that is, the "normal" variation implicit in the use of such terms as "excessive," "underdeveloped," "too low," "lagging," "unstable," et cetera. This is in no sense to say that the variation, discrepancies, and divergencies that one may observe are not "excessive" and that all is well with the best of all possible systems. Low-production families do seem to be somewhat geographically concentrated, and differences among observed states of "development" do appear odd in some way. But mere differences between incomes of different families or between average incomes of different groups of families are not in

themselves the quantities that induce in one the sense of observing "underdevelopment." It is the seeming excessiveness of the differences and the persistence in time of these seeming excessive differences that induce this conviction that something subject to choice by legislators can be improved upon. That is to say, the variation that one observes is in excess of what one might ordinarily expect of a system that is working as well as it may be made to work and is thus abnormal in this sense.

In Part III, I will raise for consideration the special kind of theoretical problem of specifying the form and magnitude of such "expected" or "normal" variation with which the observed variation is compared.

### PART III

It will be recalled that in Part I we proposed to observe and describe the physical features of the structure and development of a human population system in terms of two types of distributions. One of the distributions shows the density configuration of the loci of differentiated individual elements of the population system. This describes an arrangement of positions of things at a point in time. The other type of distribution shows the destinations of flows of things, units of value-output, et cetera, emanating from each point within the area. The two types of virtually continuous distributions specify the essential spatial structure.

In empirical studies that have been made there are at least suggestions that the distributions are characterized by forms that may be given analytical expression and that while changes of conditions over the course of time have affected the parameters of these distributions, the forms have been more or less insensitive to these changes in conditions. For example, the cumulative greater-than size distribution of cities in 1950, when plotted upon double logarithmic graph paper, appears as a straight line sloping downward at an angle of approximately forty-five degrees. Given the slope and the intercept on the vertical axis, one may quickly draw off a line from which may be read the number of cities containing more than some specified number of people. This is only an empirical "fit," and no analysis has been presented on the basis of which one may infer and rationalize a precise form of this distribution. But with an appropriate adjustment of the two parameters, the intercept and slope, the distribution for any other decennial since 1790 is given approximately by the straight line; one might hazard with some degree of confidence a straight line to describe the size distribution to come in 1970. Moreover, there are reputable persons who maintain that the size distributions of systems of central places located on other continents and in other historical periods are similarly fitted by the same form of distribution. As another example, the cumulative less-than distance distribution of carload lots of freight, when plotted on logarithmic probability paper, appears as an upward sloping straight line. And evidence suggests that this distribution is also stable over time and from place to place within the total area at the same time. Again, one may feel some degree of confidence in describing the distribution for next year by drawing such a straight line. And it may be conjectured that some stable form of dis-

tribution would be found if one had data for a large center indicating the destination of each unit of value-added produced in and immediately around the center.

In the last section of Part I, it was emphasized that if for the present we may assume that some process of development leads in the limit to stable distributions of the sort that we have described, the regularity and order so specified pertain to the population system as an entity and not to the individual elements. The individuals are not "at rest" nor ever "tending to a state of rest," being always in a continuing state of motion and change. But at any point in time, the differentiated elements seem to be distributed among arbitrary small areas in conformance with a distribution that maintains a regular and stable form. And this is the kind of situation to which *statistical equilibrium* refers.

Discussions familiar to all economists imply this latter idea. Most persons who have worked with the data expect to find a distribution of typical form when the incomes of families are arranged by size. Many would feel confident that if given as many as two or three parameters of the distribution they could construct a close approximation of the entire distribution. Now the conjecture is not infrequently expressed that if existing wealth were arbitrarily apportioned among all the families so that in an initial period each family receives the same income, and if, in subsequent periods, each family is allowed to choose its mode of using its resources and capacities, then it would only be a matter of time before incomes would again be distributed in accordance with the form which described the distribution prior to the arbitrary change. The family units would be differently placed within the distribution, but a shifting of individuals within the distribution would occur in time even though no arbitrary change were made. Whatever is done, individuals move along the scale, passing from income class to income class. The form of the distribution that is thought by many to be typical and stable will be stable only if these passages of individuals are mutually compensatory when this particular form is established. This is a limiting form, representing an equilibrium situation, if a disturbance which alters the form of the distribution renders the passages of individuals from class to class no longer compensatory but such that the initial form tends again to be restored.<sup>4</sup> This may or may not be so for our case of the income distribution, but all of us are familiar with statements in this context implying this concept of statistical equilibrium.

In the case of the movement of individuals, the apparently stable size distribution of central places is suggestive of such a statistical equilibrium. Through a continuing process of births, deaths, and movements from sub-area to sub-area within a large region, a particular form of density configuration is established. The passages of individuals from sub-area to sub-area do not alter this form. But suppose that the form of the distribution were arbitrarily altered. For example, suppose that by a decree, issued by some all-powerful authority,

<sup>4</sup> The idea is that of a stochastic process possessing ergodic properties. Cf. William Feller, "On the Theory of Stochastic Processes, with Particular Reference to Applications," *Berkeley Symposium on Mathematical Statistics and Probability*, 1949, p. 418.

individuals, supplies, and equipment were physically moved about so that all central places were made the same size with the understanding that once they were made this size, property would be divided among the individuals, and everyone then allowed to do as he chooses, subject to ordinary laws. Then these passages of individuals would tend to change this arbitrarily established distribution. And, if the original distribution does in fact conform with the idea of a state of statistical equilibrium, the arbitrarily established equal distribution of individuals among designated sub-areas would be changed, through the occurrence of undirected individual passages from sub-area to sub-area, back to the original distribution. This would not mean a re-establishment of the old cities to their former states but rather a development in time of a system of central places having the same general features of form.

This idea of statistical equilibrium is being raised here for brief discussion in order that we may consider and speculate upon the kind of analytical problem that confronts a student of the phenomenon that we have attempted to describe. If we have in what we would observe upon our map or series of maps an empirical counterpart of this concept of statistical equilibrium, our problem of analyzing the form and amount of variation represented is that of analyzing a time-dependent stochastic process. I shall now try to illustrate this notion by outlining a simple model.

One need only reflect upon the present residences of those with whom he was raised in order to be reminded of the extent of the mobility of human beings. I should think my own experience is not unusual, and in the course of my life I have moved my residence a number of times. When I recall the chance happenings that apparently determined the various moves, it seems not unreasonable for me to look upon my present residence, as well as my occupation and station, as the result of a sequence of chance occurrences. Let us consider the meaning, then, of a statement that some chance process determines the distribution of individuals over an area such as that of this nation.

We shall begin by postulating the simplest conditions imaginable. The purpose of this first step is to illustrate the idea of a random distribution of objects over an area. In our subsequent steps we shall modify the conditions of this first step in such a way that the resulting model may perhaps be regarded as having relevant features for a discussion of the real and observable phenomena.

Let us again think of a large map of this country upon which are to be placed  $r$  chips representing the family and individual units constituting the population. Let the total area be partitioned into  $N$  sub-areas and numbered 1 through  $N$ . Let a number be randomly drawn for each chip, the number to indicate the sub-area into which the chip is to be placed; and let the drawing be with replacement and such that the chance of any chip's being placed in one sub-area is the same as it is for any other.

As a result of the sequence of  $r$  draws, there will be sub-areas containing 0, 1, 2, 3, ... chips. And there would be an expected proportion of all the sub-areas which would contain exactly  $k$  chips, this proportion being given by the well-known Poisson formula.



Now suppose that this experiment were performed in each of a succession of time periods. That is, let us say that at the beginning of each time period, each of the  $r$  chips is assigned to a sub-area in accordance with the procedure outlined.

The location of each chip could be recorded for each time period, and the movements of the individual chips could be traced. And one may calculate the expected relative frequency of moves of each possible distance.

Now, the expected proportion of sub-areas containing exactly  $k$  chips may be calculated. And, if the total area were divided into several large parts—twenty or thirty or so—then, if  $N$  and  $r$  are sufficiently large, within any of these parts the proportions of places containing  $k$  chips would be given approximately by the same formula as for the total area. That is, among the various geographic divisions or regions, the ranges of heights of columns would be about the same, and the number of columns of a given height for each column of some other height would be about the same for one geographic part as for another. And thus, it is conceivable that one who only observes the results of one distribution would interpret what he sees as a “system of places,” each large column being associated with some typical number of smaller columns.

However, there would be no stability or continuity from period to period in the height of a given column. A concentration may build up in one period and vanish in the next, and moderate sized rates of growth and decline would be virtually nonexistent.

Note that in the paragraphs immediately above we have called attention to four quantitative aspects of the outcome of the process: the first has to do with the mobility of individuals constituting the population; the second with the size distribution of the columns at any point in time; the third with the spatial orientation of the columns of various sizes; and the fourth with the growth and decline of the populations of the sub-areas. A particular performance characteristic has been attributed to the process corresponding to each of these quantitative aspects.

Let us now modify the experiment. In the above illustration it was assumed that in each period the chance of a given chip's being placed in one sub-area is the same as the chance of its being placed in any other, viz.,  $1/N$ . On this assumption, the chance of a chip's remaining in a sub-area from one period to the next is this small number  $1/N$ . But empirical studies show that only about two out of ten persons move at all from one year to the next and that one of these two move no farther than from one house to another within the same county. Only about one out of ten persons move far enough to change their counties of residences, and most of these move to adjacent counties.

In the chance assignment of a chip to a sub-area, then, it would be in keeping with experience for us to use some kind of urn-and-ball schema so that the probability of a given chip's moving to a particular locality would be made to depend upon the distance of the locality from the present location of the chip. That is, with all other conditions the same, we should make the probability of a move vary inversely with the distance of the move.

But sub-areas differ, of course, in respects other than distance from a given

point. On the one hand, they differ in their natural endowments—proximity to lines of transport, quality of soils, climate, accessibility of minerals, et cetera. These would be initial differences representing a state existing at the beginning of the process. On the other hand, there are differences that result from a process of settlement. It seems plausible to assume that the presence of individuals within a sub-area increases the chance that this sub-area will be subsequently selected as a residence by other individuals. The activities of individuals are mutually dependent, and the establishment of one kind of activity within a neighborhood involves the construction and use of facilities that may be jointly used by other individuals. Initially, there are natural differences among sub-areas, and a chance of a move being made from a particular sub-area to a sub-area lying at some given distance may be greater or less than the chance of a move to some other sub-area lying at an equal distance. As moves take place from period to period, these probabilities would be altered.

In the construction of our model, let us say that an urn-and-ball schema is prepared so that there is an urn containing numbered balls for each sub-area containing any chips. The experiment would consist of drawing a ball for each chip at the end of each period, the number on the ball indicating the location of the sub-area which will contain the chip during the next period. Each sub-area, as a possible location of the chip in the subsequent period, would be represented in the urn not by one ball but by a number of balls. This number of balls would vary from sub-area to sub-area: directly with the degree of attractiveness of the natural endowments of the sub-area; directly with the number of chips presently occupying the sub-area; inversely with the distance of the sub-area from the locality for which the urn is being prepared.

Let us first consider that aspect of our model which we may suppose to correspond to the natural variation among localities is regard to physical resources. In doing so, we shall refer back to what would seem to be a paradoxical observation mentioned in Part I. On first sight one is struck by what would seem to be extreme differences among regions with respect to natural resource endowments. Also, one more or less takes for granted that the locations of human beings and their activities are strongly influenced by the locations of natural resources. Yet, the uniformities and regularities that students of the phenomenon purport to see in a spatial arrangement of a human population system appear to be largely the same whatever the time period or whatever the continent over which the population is spread. This latter observation suggests a sense of uniformity of resource endowment among the regions supporting the various systems of central places. We thus simultaneously have, on the one hand, impressive differences and, on the other, some sense of similarity. A resolution of this seeming inconsistency, I submit, is to be found (1) in attributing the differences to comparisons between sub-areas and the similarities to comparisons between relatively large regions each consisting of many sub-areas and (2) in clarifying the sense in which regions may be regarded as similarly endowed with natural resources. I shall now consider what one may suppose to be the nature of this regional uniformity of resource endowment.

Let us say that natural resources are classified into a large number of types.

Perhaps the classifications used by the Bureau of the Census or the Interior Department would not be entirely suitable, but either would serve as a beginning. For each type, it is conceivable that one might count the number of sub-areas that contain this natural resource in usable amounts. Let all the types be arrayed with respect to this number. At one extreme of this array would be a class of natural resources each member of which is found in virtually all of our sub-areas. These are the ubiquitous resources which are found nearly everywhere and which need be hauled only short distances. At the other extreme of the array would be a class of "strategic" resources; many different types of resources are included within this class, but any one of them is to be found in only a very few sub-areas. Members of this class that are widely used are hauled relatively long distances. Between these extremes there would be other classes of resources, ranging from those that are found in most places to those that are found in few. This array of types or names of resources, with respect to the number mentioned, we shall suppose to be characterized by stable properties. That is, we shall assume that about the same proportions of all named resources are to be found in the ubiquitous range of the array from period to period, and similarly for the "strategic," as well as for all other ranges. But the actual list of physical resources may well change with time, some things being regarded as valuable resources in one period that are not recognized as such in another.

Consider now how the sub-areas may vary with respect to their natural endowments. First, a sub-area may contain a smaller or larger number of different types of resources. A large proportion of sub-areas contains only the ubiquitous resources, a somewhat smaller proportion contains, in addition to the ubiquitous, some of the middle-range resources, and a very small proportion contains, in addition to the ubiquitous and middle-range, one or more of the "strategic" type of resources. Second, a sub-area may contain a higher or lower total quantity of natural resources, measured, say, by something like the f.o.b. value; a relatively large proportion of the sub-area contains small amounts of resources and a relatively small proportion contains large amounts.

These variously endowed sub-areas we suppose to be randomly distributed over the total area. For the purpose of our model, some such scheme as the following could be used for the attainment of this random distribution. Let each type or name of resource be represented by an urn containing balls, some of which are unmarked and some of which are marked, and let each of those marked have a number. There would be as many urns as there are names of resources, and for each sub-area a ball is drawn with replacement from each urn. The result of the series of drawings would determine the endowment of the sub-area in question; the occurrence of a marked ball indicating the presence of the resource and the number upon the ball the extent of the richness of the endowment of that resource. The proportion of balls in a given urn that are marked and the skew of the distribution of the numbers on the marked balls would be made in accordance with the assumptions stated in the preceding paragraphs: a very high proportion being marked in the urns representing ubiquitous types, and a very low proportion in those representing "strategic"

types, et cetera.

For each of the ten to twenty thousand sub-areas a series of drawings would be made, and the result would determine the spatial distribution of resource endowments. The process by which the distribution would be effected is such that any region consisting of a large number of sub-areas would be endowed with natural resources in about the same way as would any other, in the sense that its "average endowment" and its "variability" of sub-area endowments would be about the same as those of any other large region. This sense of similarity may be illustrated in this way: suppose one to begin with ten million black and white balls; let him randomly distribute these ten million balls among 10,000 buckets, 1,000 balls per bucket; then suppose he were to allocate randomly the 10,000 buckets among 10 or 20 groups, from 500 to 1,000 buckets per group. Think of a bucket as representing a sub-area, the number of black balls contained in a bucket as measuring the extent of its natural resource endowment, and a group of buckets as representing a region. The 10 to 20 groups would be expected to have about the same frequency distributions of buckets classified by number of black balls contained, but there would be variation, possibly extreme.

In the same sense we suppose continents and large areas within continents to be uniformly endowed. But while, in accordance with this supposition, there would be an appreciable likelihood that any large area contains resources belonging to that class any member of which is found in only a very few sub-areas, the particular member of this class that is found in a given large region is in general different from those found in other regions. That is, each region, containing a relatively large number of sub-areas, is characterized by a unique combination of "strategic" resources.

Summing up, in accordance with the above considerations and denoting by  $z$ , some measure of the quantity of resources included in the natural endowment of the  $i^{\text{th}}$  sub-area, we shall think of the distribution of  $z$  as being strongly skewed, a relatively large proportion of the sub-areas having small values for  $z$  and a relatively small proportion having large values for  $z$ . There being 10,000, more or less, sub-areas into which the total area is divided, there is this number of  $z$  values and this number of compositions of the various classes or types of resources. We think of these as being randomly distributed over the total area so that any region or zone consisting of a large number of sub-areas will contain a distribution of  $z$  values, and a composition of the different classes of resources which will be approximately those of other regions of similar size.

So much for the spatial distribution of natural endowments. Denote by  $n$ , the number of individuals located in the  $i^{\text{th}}$  sub-area. At any point in time, then, each sub-area will have a value for  $z$  and a value for  $n$ . Consider now the make-up of the urn for the chips presently located in some given sub-area. Let  $d_i$  be the distance of the  $i^{\text{th}}$  sub-area from this sub-area for which the urn is being prepared. There are now three variables:  $z$ ,  $n$ , and  $d$ . Let  $k_i$  be the number of balls placed in the urn for the  $i^{\text{th}}$  sub-area. In the preliminary construction of our model experiment, we would make  $k$  vary directly and propor-

tionately with each of two increasing functions  $f(z)$  and  $g(n)$  of the variables  $z$  and  $n$  and inversely and proportionately with an increasing function  $h(d)$  of  $d$ . That is, the  $k$  for this urn would be made to satisfy the relation  $k_i = Af(z_i)g(n_i)/h(d_i)$ . The function  $h(d)$  would be so chosen that the chance of a move to a given distance zone would vary inversely with distance. Each distance zone, if it contains a large number of sub-areas, will contain about the same relative composition of  $z$  values as that found in any other distance zone of large size. But the number of sub-areas contained within a distance zone increases proportionately with distance. Hence, if the chance of a move being made to a distance zone is to be made to vary inversely with distance, allowance must be made in  $h(d)$  to offset the effect of the increasing number of sub-areas as  $d$  increases.

We might assume that in the initial period the total area under consideration is empty of individuals. That is, in this first period the sources of migrants would lie outside of the area, as was the case in the early part of the sixteenth century, this continent being then virtually empty of Europeans. Let us say that these central places external to the area in question are represented in our model by a single urn, and let the composition of this urn be such that some plausible number of migrants would be expected to enter the initially empty area each period, and with the higher probabilities corresponding to the sub-areas on one or more of the boundaries. Subsequent to this initial period there would be sub-areas containing chips, and for any such non-empty sub-area an urn would be constructed.

The experiment would consist of the operation of this model over a long sequence of time periods. In each time period, a ball would be drawn from the appropriate urn for each chip, and the location of each chip is thus re-determined period by period. The result for the first period is a chance event depending upon the initial conditions; the initial spatial distribution of the  $z$  values and the distances corresponding to the respective sub-areas. The results for subsequent periods are chance events depending also upon the happenings in previous periods. From period to period, the  $n$  for each sub-area would be subject to change by virtue of random differences between the number of chips moving in and the number moving out. And we may allow for an additional source of change of the  $n_i$ . Some systematic birth and death process of the individual's chips may be introduced whereby the aggregate number of individuals  $r = \sum n_i$  would be subject to change apart from migration from without.

With each change of the set of  $n_i$ , the system of urns would be adjusted. Period by period the development would proceed. For any period, the size distribution of the columns could be observed; that is, the number of columns containing exactly  $x$  chips. Presumably, if the experiment were realistically designed, a particular form of this distribution would be observed after the process had continued for a long while. This would correspond to what we speak of as the process of urbanization of a population. In each period there would be observed a distribution of individual's moves classified by length. This would be descriptive of one aspect of the mobility of the population. The spatial orientation of the columns of various heights could be observed, and

again we presume that, if the right design of the experiment can be discovered, there would be observed after a number of time periods a particular pattern of this spatial distribution. That is, for any part of the total area containing a large number of sub-areas there would be expected values of the numbers of columns of the various sizes. During the course of the process, there would be an observable distribution of rates of growth or decline of the individual columns, and this distribution would presumably also tend to assume some stable form.

Some four features of structure and development have been mentioned in the preceding paragraph: (1) the size distribution of cities, (2) the spatial distribution of the cities of various sizes, (3) the distance distributions of the movements of individuals from place to place, and (4) the distribution of the growth rates of the populations of central places, sub-areas, and regions. Numbers (1) and (2) have to do with degrees of urbanization, (3) with the mobility of the population, and (4) with divergencies in rates of growth and development. We considered in Part I other aspects of variation that are not represented in this preliminary version of our model, e.g., variation among communities of different sizes with respect to types of occupations, grades, and qualities of capacities exercised by the members of the populations, and incomes of these members. With practice and experience in the operation of the model, one may hope that he could find ways of elaborating the process, composing the contents of the urns, so that these features of variation would also be included.

In the real population, each individual during his waking life apparently is engaged in a decision-making process. He consciously weighs and judges the expected consequences of alternative actions which are possible under a given system of physical and institutional constraints. But in the construction of the model experiment, we abstracted from this individual decision-making process, for this does not pose the problem with which we are now concerned. The assumption is made that the results of millions of individuals making rational choices are substantially the same as though the individual movements were chance events. And one may note in passing that this kind of assumption is also made by analysts of traffic-flow phenomena. Presumably, the driver of an automobile rationally chooses his routes, rates of speed, times of trips, et cetera. But the models constructed for explaining the regularities and uniformities in this flow phenomena are probabilistic models which include no explicit assumption regarding individual motivation and decision-making.

The problem of political economy is posed not by the necessity of an individual making a choice under a given system of constraints but rather by the necessity of a society of people making a choice among alternative systems of constraints. For a legislative body to evaluate an existing system of constraints is for the members of that body to compare, through discussion leading to practical consensus, the working properties of this system with the working properties of some altered form of the system. As indicated in Part II, what legislators literally choose is a system of rules constraining and prescribing the actions of individuals. The outcome of a people's choosing specific modi-

fications of its system is described in terms of limiting forms of distributions, examples of which would be those specifying the four aspects of variation mentioned above.

An analysis of a model of the sort outlined above would be useful to the extent that thereby one could anticipate the effect upon such distributions resulting from the adoption of an amended set of constraints. For the purpose of the analysis, the proposed changes of constraints and prescriptions would be translated into changes of the compositions of the respective urns in the model. The skill with which this translation of a modification of the constraints upon private decisions of individuals into modifications of the probabilities represented by the compositions of the urns would be developed, one may hope, through practice.

And now how is it said that any of this would contribute to an understanding of the concept of *expectations*; the concept that in Part II we argued is implicit in judgments sensed and expressed by discussants of public problems? Perhaps I am unable to express what I have in mind, but it seems clear to me. The form of the variation that is observed and pointed to by persons in the act of describing faulty performance is not itself subject to direct choice. That which is subject to being directly chosen is an amended set of constraints. Under the new set of constraints, individuals will exercise choices and make their private moves. Corresponding to the new constraints and prescriptions there will be power structures and distributions of variables descriptive of amounts of variation and of performance of the new system. To say that the observed pattern of variation is indicative of faulty performance is to imply that this observed pattern is significantly different from the pattern that would result from an adoption of a specifically amended system of constraints and prescriptions. This latter pattern of variation is the implied *expectations*.

Note well that it is not enough for one to propose less inequality of the distributions of income and wealth, less variation among communities of rates of growth or rates of unemployment, less concentration of personal power, less opportunity for persons to coerce other persons. These are more or less complicated variables in terms of which one may well observe the performance of a system, and I should think that most persons would favor lesser values of these variables. But what is actually subject to being directly chosen is not lesser or greater values of any set of variables. That which may be chosen is a new set of constraining and prescriptive rules. The analytical problem arises from one's trying to anticipate "what would tend to emerge," regarding induced changes of expected values of such variables, as a result of men's making their private moves within the new system of constraints.

---

## 2 Location Theory

William Alonso

This article tries to acquaint the reader with the theory of the location of the firm. The formal theory originated in the work of Alfred Weber and, through the contributions of later writers, developed rapidly until the 1950's, when further additions to its elegant structure seemed to bring increasing costs in terms of complexity of form and decreasing marginal returns in terms of new insights. At this time Walter Isard demonstrated its unity with the classical economics of substitution analysis, and an awareness developed of other problems of location and regions beyond the scope of this theory. Consequently scholars have turned elsewhere, and significant contributions to location theory have become rare.

In these pages are presented some of the principal insights of this branch of economic theory as it stands today. The first part of the article tries to synthesize the work of many scholars. To trace the intellectual ancestry of the various parts would be a laborious exercise; rather, a brief bibliographic note has been appended pointing to seminal and representative works. The second part seeks to clarify some of the deficiencies and limitations of this type of theory, and in doing this it points to some areas where further work would be fruitful. In particular, the theory of market areas seems to be the key to future developments. It was, in a sense, the culmination of the theory of the firm. But it was also the beginning of the theory of August Lösch, as represented by his article in this volume, and, in the form of central place theory, it constitutes the principal tool for understanding the empirical regularities that concern Brian Berry, among others. The theory of market areas is implicit too in the modern theories of regional development such as Perroux's growth poles. Thus, in a way as yet unrealized, it may be said to be the fulcrum on which turn the various spokes of the understanding of regional structure and development.

The theory of the location of the firm has been developed in the context of a free market. In recent years, however, concern has turned to national regional development. Increasingly the question of the loca-



tion of a factory is being considered as a "project" by a government agency rather than as a profit-making venture by a private corporation. It is clear that, as long as the decision turns on the project maximizing its own returns, there is little difference whether those making the decision serve a public or a private body. But the theory has little force in considering the costs and benefits accruing outside the books of the particular enterprise. Regional economics is concerned with these external or multiplicative effects within the region, and national regional planning with these effects among regions. In this sense, the theory of the location of the firm extends to project planning but antecedes regional and national spatial planning.

Because of the variety of backgrounds of those interested in regional development, and at the risk of irritating the knowledgeable, I have explained the technical terms in this article. Moreover, to avoid awkwardness of language, I have based most of the analysis on the businessman as protagonist rather than the project planner.

In essence, the firm wishes to maximize its profits. If the businessman can state clearly what factors are involved and what relations bind them, his problem is one of manipulating these variables to get the largest profits. In reality these variables are many, and some cannot be quantified. Here we shall begin with very simple problems, paring the problem down to a few essentials, and introducing complications gradually to make the theory more realistic; but some considerations will remain outside the formal theory. We shall call attention to some of these, but they are potentially infinite. For instance, a study of the New England region found a manufacturing firm in Worcester which would clearly have been better off in Boston. The reason for its location, it was discovered, was that the manufacturer's mother-in-law lived in Worcester, and his wife insisted on living in the same city. No amount of formal theory would have unearthed this reason, but formal theory could tell the manufacturer how much this cost him.

### THE PRINCIPLE OF MEDIAN LOCATION

Let us begin by considering the location of a firm which, let us say, makes and delivers bakery products. Neither the cost of making these products nor the volume of business will vary with the location of the firm. The only variable in this case is the delivery costs, so that maximizing profits is identical with minimizing delivery costs. The customers, *A*, *B*, . . . , *G*, each take one delivery a day, and are distributed along a road as shown in Figure 1. The bakery sends out a boy who can carry

only one customer's order at a time, so that he has to make one trip per customer. Where then to locate the bakery to minimize the boy's trips? The almost automatic answer would be the "average," center of gravity, or mean location. This is easily found by summing the distances from either end and dividing by the number of customers. In this case, summing from *A*, it would be  $0 + 1 + 2 + 4 + 6 + 14 + 15 = 42$ ; dividing by the number of customers or trips (7), the mean is 6 blocks to the right of *A*, at the same location as *E*. But this

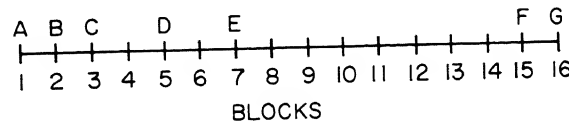


FIG. 1. Distribution of the bakery's customers.

is the wrong solution. Examine Table 1. The total distance is less if the bakery locates at *D* than if it locates at *E*. If we had gone about the problem in a systematic fashion, we should have asked: Which location minimizes the sum of the distances from the bakery to its customers? This can readily be solved by elementary calculus. In fact, however, we might have recognized that the point on a distribution along a line at which the total distance to all other points is minimized

TABLE I  
TOTAL TRIPS ACCORDING TO LOCATION OF THE BAKERY AT E OR D IN FIGURE 1

| Customer       | Distance from<br>location at E | Distance from<br>location at D |
|----------------|--------------------------------|--------------------------------|
| <i>A</i>       | 6                              | 4                              |
| <i>B</i>       | 5                              | 3                              |
| <i>C</i>       | 4                              | 2                              |
| <i>D</i>       | 2                              | 0                              |
| <i>E</i>       | 0                              | 2                              |
| <i>F</i>       | 8                              | 10                             |
| <i>G</i>       | 9                              | 11                             |
| Total Distance | 34                             | 32                             |

is the median (that is to say, the point at which there are as many points to one side as to the other). The median in this case is *D*. The mean or center of gravity, on the other hand, minimizes the sum of the squares of the distances, and therefore is irrelevant for our purposes.

This simple example is a very enlightening one. We would not often

meet a bakery in these precise conditions, but the logic often applies to other enterprises. For instance, a factory which has shipping costs proportional to the weight of the shipments and the distance shipped would benefit from locating at the median location unless there were strong reasons to the contrary. Thus, a firm selling 200 units in one city, 300 in a second, and 550 in a third would have its median location in the third city. Since the median of the distribution of customers will tend to be in large cities, this is one of the reasons why big cities tend to grow bigger.

### COMPETITION ALONG A LINE

It is a truism that what may be good for someone may be disadvantageous for another, but this is often forgotten. It is important, therefore, to make clear whose point of view we are considering when we say that a location is optimal.

Imagine a long beach, with people evenly distributed along its length. Each person on the beach buys one ice-cream cone, and will walk as far as necessary to get it, though he will naturally prefer to walk the shortest possible distance. If there is a single vendor of ice cream, he will not care where he locates since every customer on the beach will walk as far as necessary to buy his cone. Every customer, however, would prefer to minimize his walk by having the vendor as near to him as possible. A third point of view might be that of a public official who wants to minimize the total amount of walking for the general benefit. As shown in our first example, this total will be minimized at the median location, in this case the midpoint of the beach.

Consider now the same problem with two vendors, *A* and *B*, who are at two locations as in the first stage of Figure 2. Vendor *A* will sell to all the customers to his left, and Vendor *B* to all those to his own right; of the customers between the two, the left half will go to *A* and the right to *B*. But *A*, after examining the situation, decides that by moving to the right he can take away many of *B*'s customers without losing any of his own (second stage). Vendor *B* then decides to hop over to the left of *A* (third stage). It can easily be seen that the final stage will have both *A* and *B* together at the center of the beach, each selling to half the customers. Neither *A* nor *B* will then be able to increase the number of his customers by moving, and the situation will be stable.

Free competition in this case will result in both vendors joining at the middle of the beach. Since people are evenly distributed along the length of the beach, the average distance walked will be one fourth of

the length. A public official might point out, however, that this average distance is unnecessarily high. If the two vendors were located at the quarter points, as in the "planned location" of Figure 2, the average

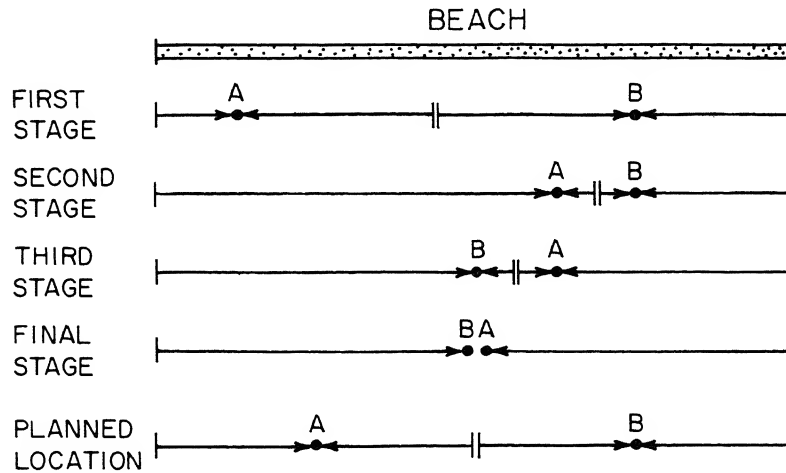


FIG. 2. Locations of vendors on the beach.

distance walked would be reduced by half, while both vendors would still enjoy the same sales. To obtain the benefits of this solution it is only necessary to assure each vendor that the other will not start moving in on him.

There are two lessons to be learned from this. On the one hand, we have seen once again a tendency toward concentration, which may be interpreted as another indication of the reasons for the development of centers of human activity. On the other, we have seen that the solution of free competition may differ from that of the public interest. This is not to say that the results of private initiative need be in conflict with the interests of the community. In fact, it will be seen that theory indicates they usually coincide. But this coincidence of interests is something to be proved in each case, rather than something to be taken for granted.

#### THE FIRM WITH ONE MARKET AND ONE RAW MATERIAL

Let us consider an activity that uses only one material and sells all of its product at one market. Such a firm might use sheets of steel from a steel plant at *M* in Figure 3 as its raw material, and bend them into

boxes which it sells in a city at  $C$ . For simplicity say that the costs of production are the same everywhere, so that the firm's only consideration will be to minimize its *total transport costs*. These consist of *assembly costs* of bringing the steel from  $M$  to the factory, and *distribution costs* of sending the boxes from the factory to  $C$ . Let  $T$  be the

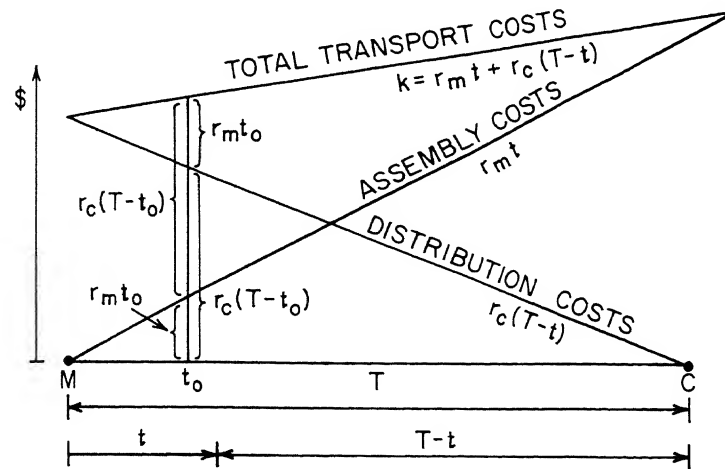


FIG. 3. Transport costs of a firm with a single market and a single material; transport costs proportional to distance.

distance from  $M$  to  $C$ , and  $t$  the distance from  $M$  to the box factory. The distance from the factory to  $C$  is the remainder of  $T$ , or  $(T - t)$ . Now, if the cost per mile of carrying enough steel to build one box is  $r_m$ , assembly costs per unit will be  $r_m t$ ; and if the cost of carrying one box is  $r_c$  per mile, distribution costs per unit will be  $r_c (T - t)$ . The total transport costs, which we represent by  $k$ , are the sum of these:

$$k = r_m t + r_c (T - t) \quad (1)$$

The firm will locate at that value of  $t$  that minimizes  $k$ .

Examine Figure 3, where assembly, distribution, and transport costs are shown for one case. The curve of total transport costs is the sum of the other two, as illustrated for location  $t_0$ . In the case illustrated it can be seen that transport costs will be least when the box factory is at  $M$ , where  $t = 0$ . As the diagram is drawn, the curve of assembly costs is steeper than that of distribution costs, meaning that it is more expensive to move steel than boxes; or, to put it another way, that the transport rate for steel ( $r_m$ ) is greater than that for boxes ( $r_c$ ). We may

rewrite Equation 1 as  $k = (r_m - r_c)t + r_cT$  without changing its meaning. From this form of the equation it can be seen that when  $r_m$  is greater than  $r_c$ , as in Figure 3, the firm will want to keep  $t$  as small as possible (that is, locate at  $M$ , where  $t = 0$ ). But when  $r_c$  is greater, the coefficient of  $t$  will be negative, and the firm will want to locate at the maximum  $t$  (that is, at  $C$ , where  $t = T$ ). Finally, if the costs of moving boxes or steel are equal, so that  $r_m = r_c$ , the coefficient of  $t$  will be zero and transport costs will be  $r_cT$  wherever the plant locates. The plant may then locate at  $M$ , at  $C$ , or at any point in between.

### THE STRUCTURE OF TRANSPORT COSTS

Although the costs of transportation do increase with distance, it is not accurate to say, as we have been saying, that they increase in direct proportion. In the first place, there are terminal costs: the costs of putting things on a truck or train, and of taking them off, the costs of packaging and certain paper work. These in general will not vary with distance. Therefore, transport costs are better represented by an expression such as  $s_m + r_mt$ , where  $s_m$  are the terminal costs,  $r_m$  is the rate per mile, and  $t$  the number of miles. Thus, above we used the expression  $r_mt$ , which results in a straight line passing through the origin, such as  $A$  in Figure 4. Introducing terminal costs will add  $s_m$  all

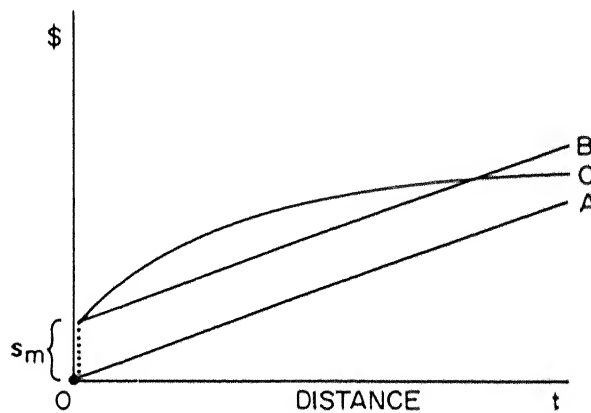


FIG. 4. Transport costs: (A) proportional to distance, (B) considering terminal costs, (C) considering decreasing marginal costs.

along to line  $A$ , resulting in line  $B$ . Of course, at  $t = 0$  nothing has been moved and costs will be zero. This is indicated by the dot at the origin in Figure 4.

Another realistic refinement considers that the rate per mile is lower for longer hauls. This is common practice in the transportation industry as in most others: costs are lower when buying in bulk, it is cheaper to rent by the month than by the week, and so on. This results in a flattening of the slope of the curve of transport costs with increasing distance, reflecting the lower rates. In practical terms, it is cheaper to make one 1000-mile shipment than two 500-mile ones. The curvature of the curve of transport costs is increased by the variety of carriers: ship, train, truck, pipe, and so on. Usually trucks have lower terminal costs but higher per-mile costs than trains, as do trains with respect to ships. Figure 5 shows the relation of transport costs to distance when alternative carriers are considered. The shipper will choose the carrier with the lowest costs for a particular distance, so that his effective curve of transport costs will be the heavy line in Figure 5, which is more curved than that of any one carrier.

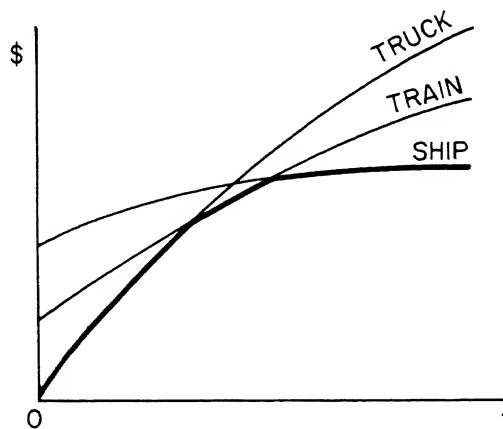


FIG. 5. Transport costs considering alternative carriers.

Terminal costs and the curvature of transport costs reinforce the attractiveness of end-point locations, such as  $M$  or  $C$  in the example of Figure 3. In Figure 6 the case of identical transport costs for steel and boxes is re-examined under this more realistic structure of transport costs. When we considered the case without regard for terminal costs or the economies of longer hauls, we concluded that when the rates are the same for assembly and distribution costs the firm would locate at  $M$  or  $C$  or at any intermediate point. But now, the curvature of the curves of assembly and distribution costs leads to a curvature of the curve of total transport costs, with the midway point the costliest although assembly and distribution costs are symmetrical. The economy

of long hauls points to location at either end since the continuous curve of total costs reaches as low as  $OY$ . But by locating at either  $M$  or  $C$  the firm can save the terminal costs on either steel or boxes, and have only  $OX$  transport costs. Even without considering the curvature, the

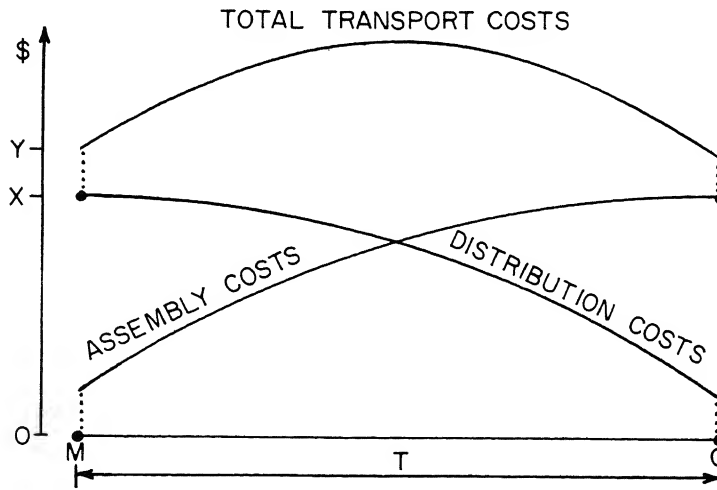


FIG. 6. Transport costs of a firm with a single market and a single material, considering terminal costs and decreasing marginal costs.

elimination of one set of terminal costs would lead to location either at the source of the material or at the market. Although the conclusion requires some reservations, this is part of the explanation for the spatial concentration as opposed to the dispersion of industry.

#### TRANSSHIPMENT POINTS: THE IMPORTANCE OF PORTS

One special case deserves attention since it accounts in large measure for the existence of many great cities of the world such as New York, London, and Buenos Aires. This is the case of points of transshipment, of which a seaport is a prime example. At these points things brought in by water must be taken off ships and put on trucks or railroads, and vice versa. This provides an excellent opportunity to process materials as they are being taken off one carrier and before they are put onto another. For instance, the American Midwest ships wheat to Buffalo by water. There it is taken off the ships, milled into flour, and the flour shipped by train to the bakeries of the Eastern markets. Or petroleum is brought by ship to New York, there refined, and the petro-



leum products are sent to other cities. Or cattle is brought from the interior by rail to Buenos Aires, there slaughtered, tinned or frozen, and shipped to foreign markets.

A diagrammatic analysis of the advantages of a transshipment point is presented in Figure 7 for a one-material, one-market concern such as our steel-box manufacturer. Let us say that steel is produced at *M*, which is separated by a sea from *B*, from which there is rail connection

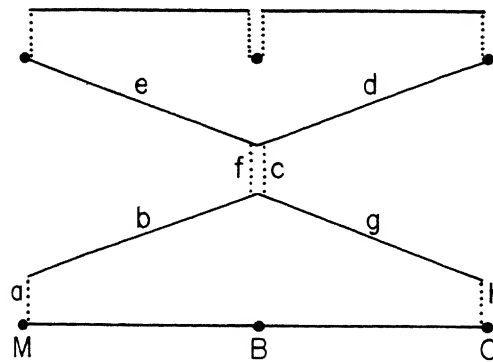


FIG. 7. Transport costs involving a transshipment point.

to the market at *C*. The curve of assembly costs is  $a - b - c - d$ , where *a* indicates the costs of putting the steel on a ship and taking it off, *b* the costs of moving the steel across the sea from *M* to *B*, *c* the costs of putting it on and taking it off the train, and *d* the costs of moving it from *B* to *C*. The curve of distribution costs is  $e - f - g - h$ , where *e* are the costs of moving the boxes from *M* to *B* across the sea, *f* are the costs of putting them on and taking them off the ship, *g* are the costs of moving them from *B* to *C*, and *h* the costs of putting them on and taking them off the train. The top curve represents the total costs of transportation, and is the sum of the other two curves. Because of terminal and transshipment costs it has three low points: if the plant is located at *M*, the costs will be  $e + f + g + h$ ; if it is located at *B*, the costs will be  $a + b + g + h$ ; and if it is located at *C*, they will be  $a + b + c + d$ . In Figure 7 all three are shown as having the same total transport costs, but which will be best will depend on the particular values of the components in each case. Ports owe their growth to the fact that they often turn out to be the best location.

The existence of a transshipment point clearly depends on the technology and development of transportation. Thus, some believe that the

development of the St. Lawrence Seaway, which permits direct shipment by water from the Midwest to world markets, may affect adversely cities such as Buffalo and Montreal, which have been transshipment points. There have been instructive instances of artificial transshipment points at locations where railroads of different gauges meet, or where two railroad networks serving a city have purposely refused to interconnect. Even such trivial breaks in transportation can bring about local development. Within cities, commuter railroad and subway stations represent transshipment points for those who use them, and generally foster development of local centers of activity. Looking into the future, one may well speculate the possible effects of the development of craft that travel on a cushion of air, able to move indifferently over land or water. It would seem that should such craft prove to be economical, they would seriously threaten the age-old pre-eminence of ports.

#### LOCATION OF INDUSTRY WITH MANY RAW MATERIALS

In the analysis of the location of firms with one market and one source of raw material we used diagrams (such as that in Figure 3) of only two dimensions: the horizontal for the distance, and the vertical for the costs of transportation. But when we consider distances between three or more locations, a one-dimensional straight line is not enough and we need a map, which uses up the two dimensions available on a page. Now the costs of transportation require a third dimension. Although difficult, it would be possible to work with three-dimensional models to analyze these cases; there is, however, a simpler way of handling them. In the upper part of Figure 8 the transport costs from some point  $A$  are shown much as in Figure 6, except that we are considering the possibility of movement in both directions from  $A$ . In reality we have in mind movement in every direction from  $A$ , so that the transport costs would look like a windblown umbrella as shown in the side-diagram, where the stem represents the terminal costs, and the umbrella itself the movement costs. The lower part of Figure 8 shows the costs seen from above as in a map. The \$2 level of transport costs would be a circle around  $A$ , the \$3 level a larger concentric circle, and so forth. The meaning of each circle is that one unit of whatever is being shipped from  $A$  can be carried to any point in the circle at that cost. At  $A$  itself, of course, the transport costs are zero.

Consider now the location of a firm which uses two raw materials,  $M_1$  and  $M_2$ , and sells its products at a city  $C$ . It is necessary to stand-

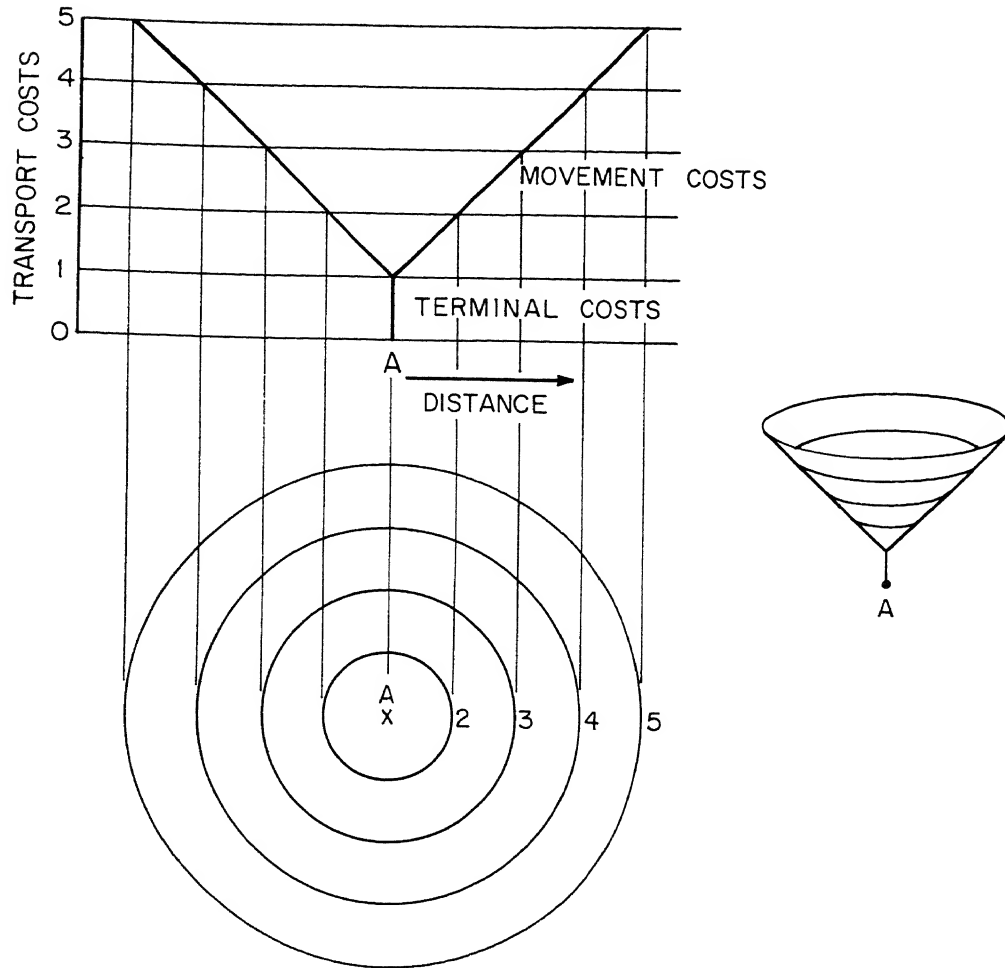
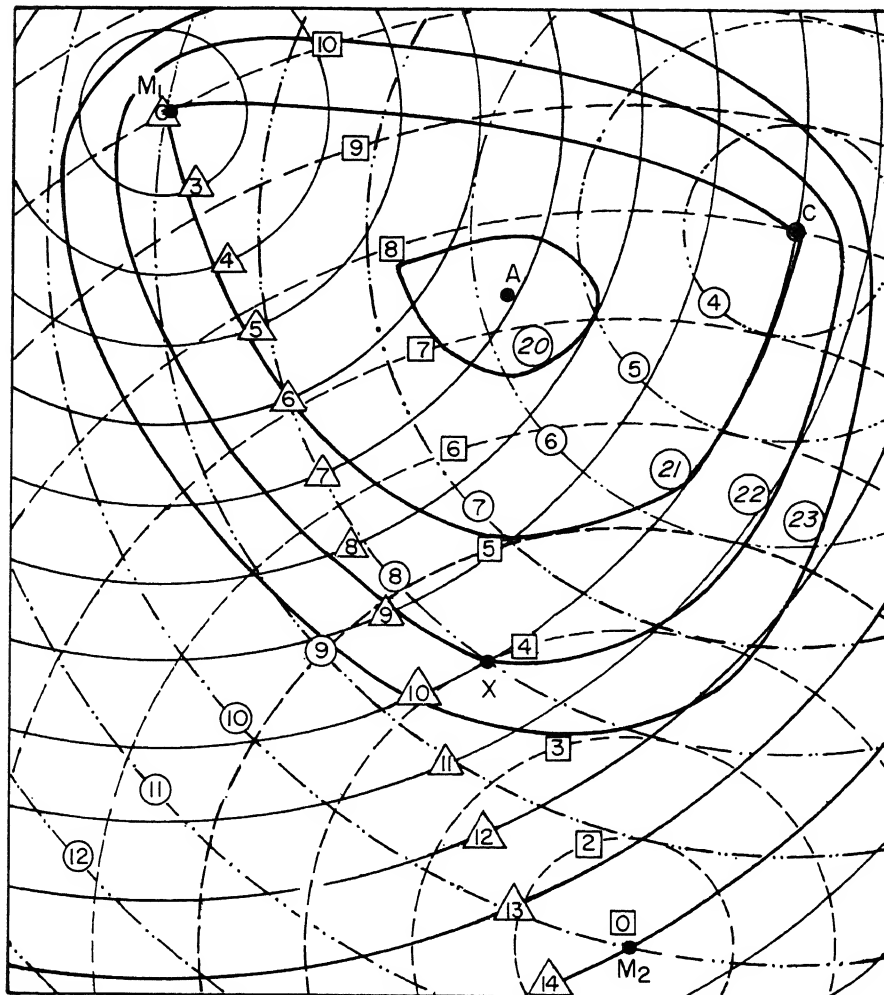


FIG. 8. Transport costs: three diagrammatic representations.

ardize the quantities per unit of product; let us say that for one unit of product we need two tons of  $M_1$  and one of  $M_2$ . We find that terminal costs per ton are \$1.00 for  $M_1$  and for  $M_2$ , so that terminal costs *per unit of product* are \$2 for  $M_1$  and \$1 for  $M_2$ . Movement costs per ton are \$0.67 per hundred miles for  $M_1$  and \$1 for  $M_2$ . Movement costs per unit of product, therefore, will be \$1.34 for  $M_1$  and \$1 for  $M_2$ . The terminal costs for the product are \$3, and the movement costs \$1 per hundred miles.



- △— Assembly costs from  $M_1$  (isotims)
- -□- - Assembly costs from  $M_2$  (isotims)
- ...○... Distribution costs to  $C$  (isotims)
- Total transport costs (isodapanes)

FIG. 9. Isotims and isodapanes for a firm with two materials and one market.

We may now carry out the analysis as in Figure 9. We draw around  $M_1$  the transport costs for the two tons needed per unit of product, shown by the thin continuous lines in Figure 9. These curves are called *isotims*. Similarly, we draw the transport costs of moving the necessary quantity of  $M_2$ , shown by the dashed circles. And finally, we draw the isotims for the product, centered around the market at  $C$ , shown by the dot-and-dash lines. The total transport costs at any point will be the sum of the isotims; for instance, at point  $X$  the costs of bringing two tons of  $M_1$  are \$10, the cost of bringing one ton of  $M_2$  are \$4, and the cost of delivering the product to  $C$  is \$8. Total transport costs, then, are  $10 + 4 + 8 = 22$ . As total transport costs are calculated over the map, points with the same total costs may be joined. The resulting lines, shown by the heavy solid curves in Figure 9 are called *isodapanes*, and constitute a mapping of total transport costs.

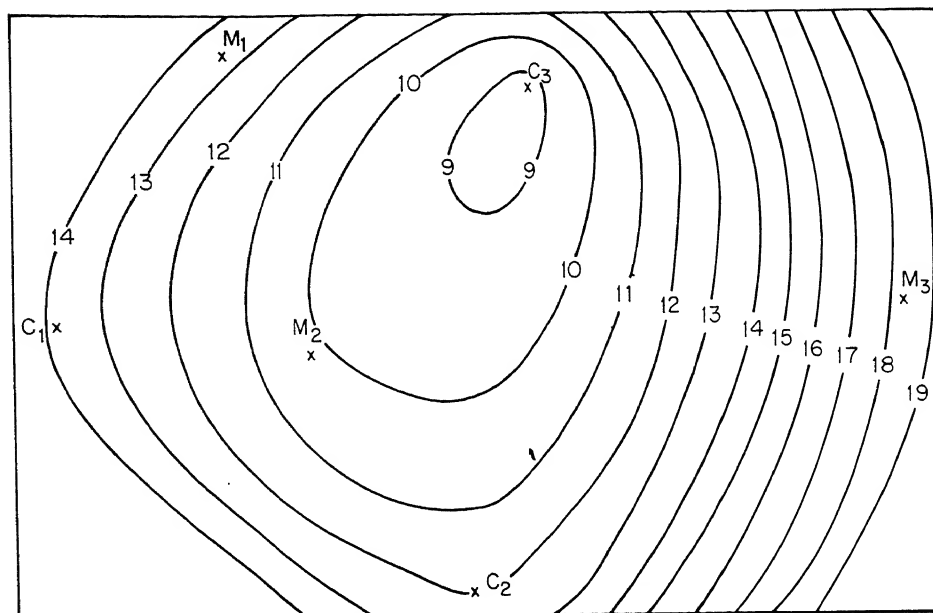


FIG. 10. Isotims of combined distribution costs.

To locate the plant, we want to find the point of least transport costs. The isodapanes in Figure 9 show a low point in transport costs at point  $A$  within the \$20 isodapane. In many cases the true minimum may be at such an intermediate point; that is to say, at a point which is neither one of the sources of materials nor the market. In this case, however, the intermediate location  $A$  is only a relative minimum. Location at

$M_1$  results in only \$19 transport costs (\$10 from  $M_2$  and \$9 to  $C$ ); and location at  $C$  in \$18 total costs (\$10 from  $M_1$  and \$8 from  $M_2$ ). The best location, therefore, would be at  $C$ . The minimum found by isodapane mapping should be checked against location at the sources of materials or at the markets to insure that the true minimum is found.

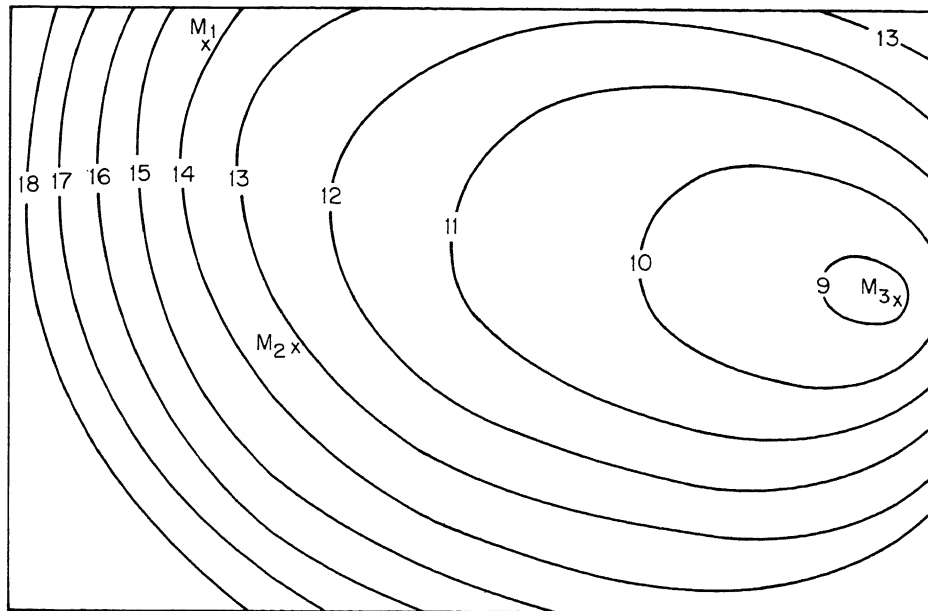


FIG. 11. Isotims of combined assembly costs.

By this method we may consider a problem involving any number of points. Figures 10, 11, and 12 deal with an industry with three markets,  $C_1$ ,  $C_2$ , and  $C_3$ , and three sources of materials,  $M_1$ ,  $M_2$ , and  $M_3$ . For simplicity, let us say that terminal costs are insignificant, although they could be considered just as they were in the discussion of Figures 4, 6, and 7.

The firm sells 20 per cent of its products at  $C_1$ , 30 per cent at  $C_2$ , and 50 per cent at  $C_3$ . In this case we may observe in advance, from the rule of median location, that the minimum of these distribution costs must be at  $C_3$ . However, a full mapping of the isotims of distribution costs will be necessary to combine with the assembly costs to find total transport costs. Since the proportions shipped to each market are known, we may draw a set of isotims for each market. If the transport rate per unit of product is \$4 per hundred miles, we may consider that 0.20 of that unit will be shipped to  $C_1$ , at a cost of \$0.80 per hundred miles; that 0.30 units will be shipped to  $C_2$  at a cost of \$1.20

per hundred miles, and 0.50 units to  $C_3$  at a cost of \$2.00 per hundred miles. If we were considering terminal costs, they too would be divided proportionally. On this basis a set of isotims may be drawn around each

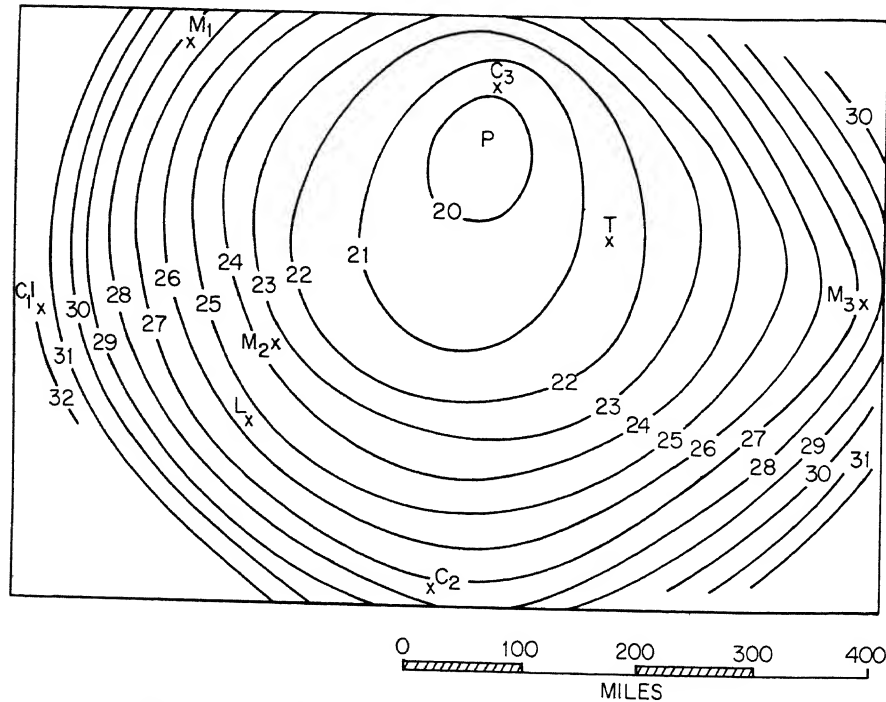


FIG. 12. Isodapanes: sum of isotims in Figures 10 and 11.

of the markets, and they may be summed (in the same way as in Figure 9) to obtain a set of distribution costs isotims. The resulting distribution costs isotims are shown in Figure 10. Note that the lowest point is indeed at  $C_3$ , where distribution costs are \$8.65 per unit.

Similarly, we may draw the isotims for each of the raw materials and sum them to obtain an isotim mapping of assembly costs, as has been done in Figure 11. The basis of this mapping is as follows:

|  | $M_1$  | $M_2$ | $M_3$ |
|--|--------|-------|-------|
| Units of material per unit of product                    | 2      | 1     | 1     |
| Rate per unit of material per 100 miles                  | \$0.50 | 0.50  | 2.00  |
| Rate per units of material necessary per unit of product | \$1.00 | 0.50  | 2.00  |

Now, total transport costs (isodapane mapping) may be obtained by summing the isotims of distribution costs (Figure 10) to the isotims of assembly costs (Figure 11). The result is Figure 12. Minimum total transport costs will be at point  $P$ , where they will be \$19.70 per unit of product, and the firm would locate there. Clearly, this minimum must always be within the polygon whose vertices are the locations of the markets and materials.

It would have been possible, of course, to have added the six sets of isotims (from  $M_1, M_2, M_3$ , to  $C_1, C_2, C_3$ ) at the same time, but having so many lines in one map is confusing, and it is easier to do it by parts.

### PRODUCTION COST DIFFERENTIALS

Suppose that at point  $L$  in Figure 12 there is a city with surplus labor, so that wages are lower than elsewhere. The manufacturer wants to know if he should locate there rather than at  $P$ . He would calculate the savings per unit of product that cheaper labor would imply, considering both the wage rates and the efficiency of that labor. If the saving is, let us say, \$10,  $L$  would be the best location, since total transport costs at  $L$  are shown in the isodopane mapping as \$25.50, as compared with \$19.70 at  $P$ . The extra transport costs, then, are \$5.80, leaving a net savings of \$4.20. We could consider, similarly, another point such as  $T$ , where there is a tax saving of \$1 per unit of product (would  $T$  be better than  $P$ ?, better than  $L$ ?), or any other point at which special conditions obtain, such as special climatic conditions, association with other activities, and so on.

### MARKET AREAS

If a firm needs a certain raw material that may come from either of two sources, the choice of one source or the other will depend on the location of the firm. But to decide the location of the firm we must know which of the two sets of isotims to consider. To do this we delimit the areas best supplied by each of the alternative sources, and consider only the isotims of the preferred source within its market area. In other words, we construct an isotim mapping for the material, rather than for the sources as such.

In Figure 13 two alternative sources  $M$  and  $M'$  of one material are considered. In the upper part of the figure are shown the delivered costs from each of the two sources. The stems are the production costs for the material plus the terminal costs, while the gradients are



the costs of moving the material over space. It can be seen that to the left of  $A$ ,  $M$  can deliver more cheaply, while to the right of  $A$ ,  $M'$  has the advantage. In the bottom part of the figure, the analysis is carried out by curves similar to isotims except that the cost of production as

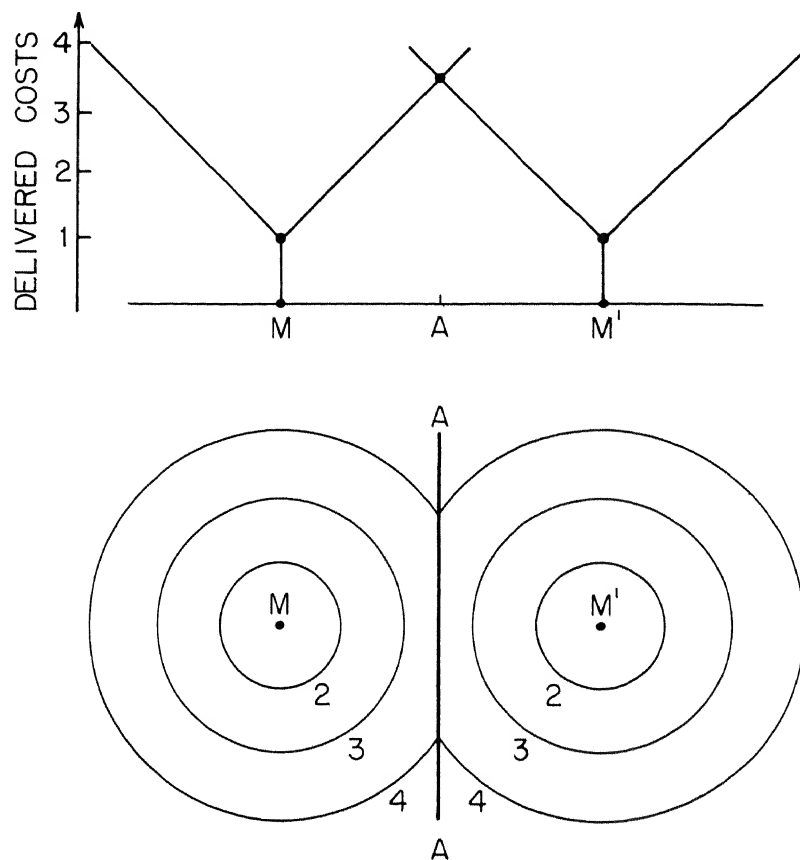


FIG. 13. Market areas: identical production and transport costs.

well as that of transportation is considered at every point. The line  $A-A$  (the perpendicular bisector of the line  $M-M'$ ) is the market boundary between  $M$  and  $M'$ . In constructing the isodapane mapping we would use isotims centered about  $M$  to the left of  $A-A$ , and isotims centered about  $M'$  to the right of it.

In Figure 14 another case is considered, where production costs are greater at  $M'$  than at  $M$ , but transport rates are the same. The resulting market boundary is an open hypercircle  $A-A$  (similar to a hyperbola), as shown in the figure. In Figure 15 a case is shown for

which transport rates are higher for  $M'$  than for  $M$ . The market area of  $M'$  will be that bounded by the closed hypercircle  $A-A$ . This situation might arise, for instance, if we were considering coal mines, and

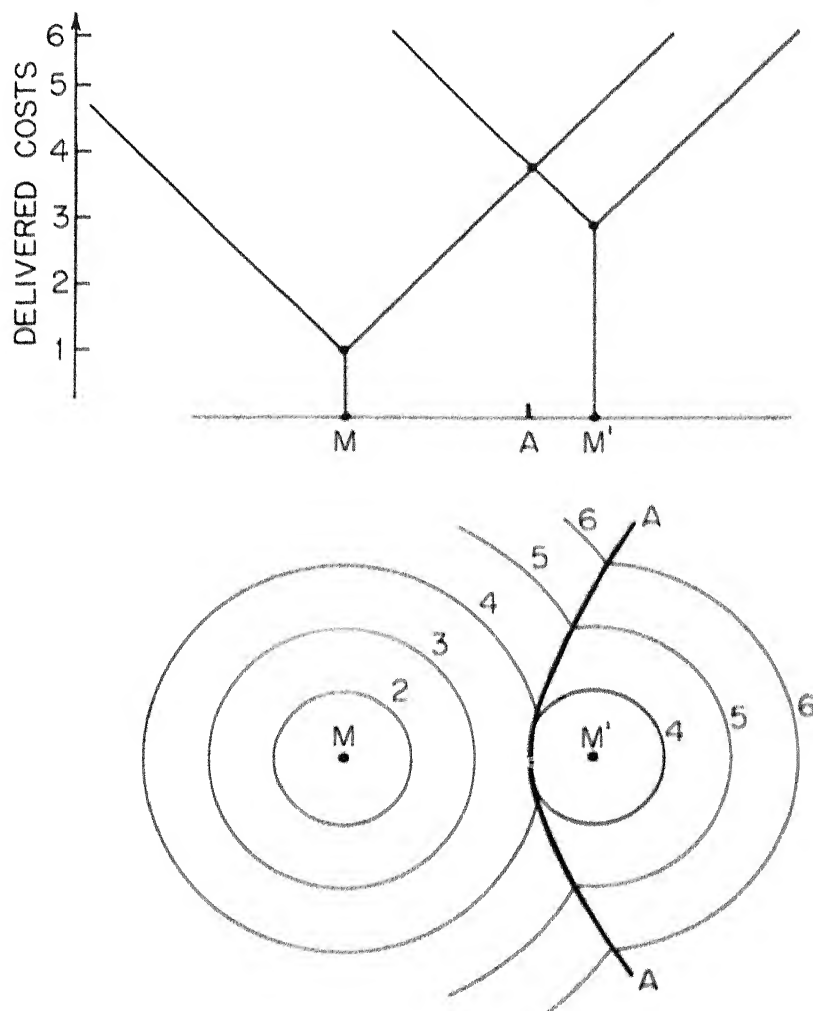


FIG. 14. Market areas: different production costs and identical transport rates.

the coal produced at  $M'$  were of inferior quality, so that greater quantities per unit of product are necessary.

This analysis holds as well for determination of the markets of firms of known location if their pricing policy is such that the customer bears

the transport costs. However, if the producers charge the same "list" price everywhere, their markets could not be determined in this way. It would depend on the policy of the producer with respect to how far

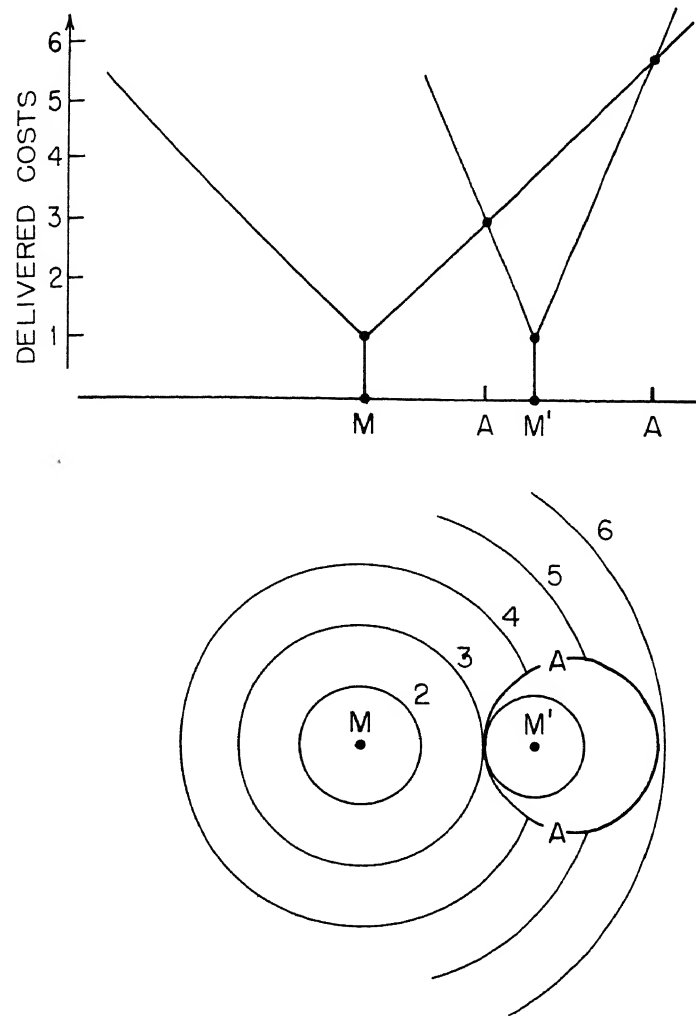


FIG. 15. Market areas: different transport rates.

he would be willing to ship his products when he himself absorbs the transport costs. Moreover, in some cases firms may engage in quite elaborate price wars to capture certain markets, or may be able to sell in certain areas even at higher prices because of advertising.

## SOME REALISTIC COMPLICATIONS

In the preceding illustrations isotims have been shown as evenly spaced concentric circles. The economies of longer hauls, which were reflected in a curvature of the transport gradient in Figure 4, will be reflected in increasingly wider spacing of isotims with increasing distance from the source.

There are other realistic features that may be considered by the use of isotims. We have been assuming that transportation is equally possible in all directions. But if we realize that there exist roads and railroads only in some directions, we see that, rather than perfect circles, isotims will take forms more similar to starfish, with arms extending along the transport routes, as in Figure 16, where  $A-A$  and

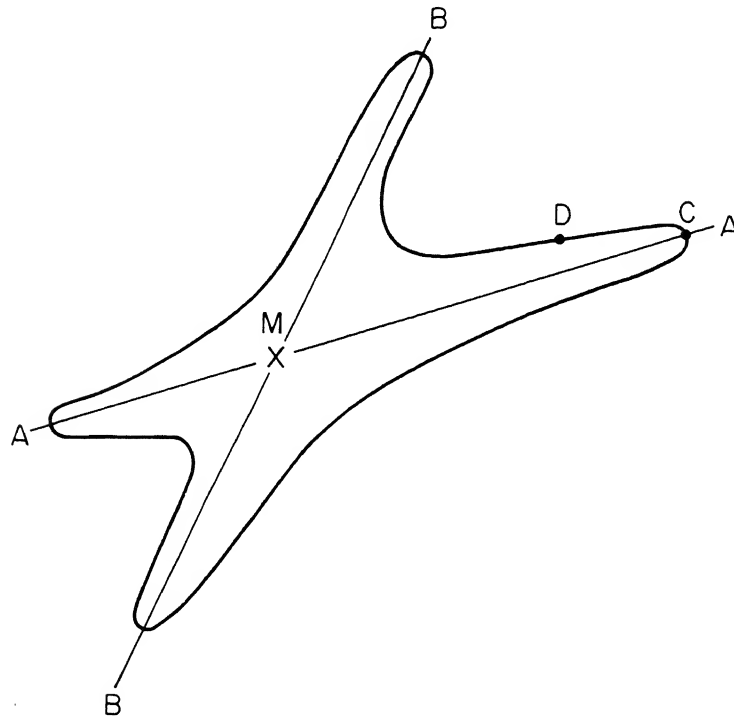


FIG. 16. Isotim considering transport network.

$B-B$  are highways or railroads and a single isotim is shown. Although point  $D$  is nearer than  $C$  to  $M$ , it is as expensive to reach, since part of the travel must be over routes inferior to  $A-A$ . Similarly, we could

consider the effects of intervening seas or lakes, mountains, or tariff barriers, routes with fixed stops that may require some backtracking to reach certain points, and so on. All of these will complicate the geometry of the isotims, but not their logic: an isotim is a curve over a map that joins points of equal transport costs.

### SOME TERMS AND SHORTCUTS

The analysis thus far could to some extent be paralleled by a device of strings, pulleys, and weights where, disregarding terminal costs, the weight at each material or market point is proportional to the weight per unit of product times the relevant transport rate, and all the strings are tied together at one knot. This knot will come to rest at the optimal transport location. More formally, the pull at each point is  $w_i r_i$ , where  $w_i$  is the amount of  $i$  necessary per unit of product, and  $r_i$  is the cost of moving one unit of  $i$  a given distance. For the product itself,  $w_i$  is unity if there is only one market, and the proper fraction if there is more than one market. The quantity  $w_i r_i$  is called the *ideal weight*. Clearly, if one of these ideal weights is greater than the sum of all the others, it will pull the knot right to its pulley. Such an ideal weight is called a *dominant weight*; represented formally,  $w_j r_j$  is the dominant weight when  $w_j r_j \geq \sum_i w_i r_i$ ,  $i \neq j$ . The significance of the dominant weight is that if there is a location with the least transport costs, it will be at the source of the dominant material (or at the destination, in the case of the product). Therefore, it is not necessary to do the isotim-isodopane analysis, since we know the answer. It will be noted, of course, that this is no more than the principle of median location in another form.

But what if there is no dominant weight? If we disregard terminal costs, this means that the point with least transport costs will be within the polygon formed by drawing the lines between the various sources of materials and markets, but not at either a source or a market. In other words, the full analysis must be carried out, and because of terminal costs the low point of transport costs within the polygon must be compared with the costs at the various sources and markets. Thus, the best location may, after all, be at one of the terminal points.

In the production of some goods, much of the materials is wasted or expended: dirt is washed from ore, material is cut into shape, chemicals are used up. In such cases, of course, the product will weigh less than the materials that go into it. Such industries are called *weight-losing* and are often *material-oriented*. The classical example of

this type was the iron and steel industry, which burned great quantities of coal per ton of product, and therefore was usually found in coal-producing regions. Modern technology, however, has greatly reduced the quantities of coal used, and reduced the material-orientation of the industry.

Conversely, there are cases in which the product weighs more than the materials that go into it. This strange-seeming phenomenon is quite frequent, and may stem from either of two sources. The freight rate of the product may be very high, because of refrigeration, fragility, great bulk, or other reasons, so that the ideal weight of the product is very high. This is the case with some glass products and precision machinery. The other common reason is the use of an ubiquitous material, such as water or air, which is available everywhere and thus does not need to be transported. Such ubiquitous materials do not enter into the calculations of dominant weights, and this type of industry is called weight-gaining and often is market-oriented. A prime example of this type of industry is beer production, which locates its plants at the major markets. But calling an industry material-orientated or market-orientated only tells us of a tendency. These terms are verbal shortcuts, similar to the concept of dominant weight, but less precise. For any particular firm, it is necessary to do a full analysis to be certain of its best location.

We have been considering industries which find transportation costs of paramount importance in selecting a location, that is, *transport-oriented industries*. Typically these are industries with a high bulk-to-value ratio for materials or products; or to be more precise, with high ideal weight-to-value ratios. For other industries, especially those with a high value-to-bulk ratio, other considerations may be more important. Thus, textiles are usually attracted to places with abundant cheap labor and are called *labor-oriented*, and aluminum is attracted to cheap electricity and is called *power-oriented*. Another type of orientation is emerging that might be called *amenity-orientation*. Research industries, such as electronics, have small transport costs but need very specialized scientists and engineers. To attract them these firms often locate where there are climatic or cultural advantages. The Worcester manufacturer mentioned at the beginning of this chapter was, in a sense, amenity-oriented. It will be noted, however, that these various orientations are really instances of the production cost differentials, and that they can be integrated with the analysis of transport costs.

Industries that have no strong locational preferences, and particularly industries that are not transport-oriented, are often called *foot-loose*, and

there is good reason to believe that technical developments are making more industries foot-loose. In the first place, in the long run, transportation tends to become cheaper, quicker, and more efficient, lessening transport-orientation. Second, production processes tend to become more efficient, requiring less materials per unit of product, thus reducing total transport costs, and, incidentally, increasing market orientation. Finally, products themselves are improved and made more efficient, so that they do more of a job per unit of weight.

### CONCENTRATION AND DISPERSION OF INDUSTRY

Does increasing foot-looseness mean that industrial activities will become increasingly dispersed? Probably not. Foot-looseness simply means that transport costs are relatively less important, not that one place is as good as any other. Industries may now be attracted to areas of good weather, either because it is important to their operations (as in the case of the aircraft industry), or because it will be attractive to their workers (as in the case of some research industries). Or they may be attracted to special site advantages, or to cheap labor, or, perhaps most important, to contacts. These contacts are infinitely varied in their forms. They may be managerial exchanges, where vital information is exchanged casually over lunch, or close supplier-customer coordination, or the chance remark that discloses an unsuspected opportunity, or the shoptalk of technical people that stimulates new ideas. The importance of contacts will probably increase the attraction of large urban centers for many industries, and lead to further concentration.

Concentration in cities does not necessarily mean concentration in downtown. We are using the word city in its technical sense of the geographic extent of the area of homes and work places of an urban concentration. Thus, by "city" we may mean a metropolitan area, central city and suburbs. Since World War II, much of the industry in metropolitan areas in the United States has moved to the suburbs. The principal reason for this movement is that plants need more land because one-story buildings are better suited to today's production processes and because vast areas are needed for workers' parking. Since much more land is needed, plants go to the suburbs, where it is cheaper. The increasing shift from railroads to trucks has helped to make this possible, since trucks are more versatile and can better service industry dispersed over a metropolitan area. This dispersion or decentralization *within* metropolitan areas has sometimes been confused with a national dispersion.

Although it is hard to generalize about industrial location patterns within cities (both because we know little about it and because much of what we know requires very cumbersome technical description), one frequent pattern or life cycle of industries has become relatively recognizable. It is similar to the cycle of residential migration for American families: young couples or those recently arrived from rural areas live at first in small apartments near the center of the city. As the families grow or the rural immigrants become adapted, they tend to move to newer houses in the suburbs. Similarly, many new industries get their start in the old buildings near the center of the city, where they can rent space relatively inexpensively. When the firm becomes successful and needs more space, it frequently builds its new plant in the suburbs. In our sense of the word, this is a move to another part of the city, but not a movement out of the city.

#### HOW IS IT DONE IN PRACTICE?

We have been discussing the logic of the location of industries. But the question may be raised whether businessmen or project planners do in fact follow the methods outlined here. Many, of course, do not. In many cases the decision is made almost by whim: the businessman may simply operate his business in the town in which he was born and raised, or in a city that has persuaded him with a clever promotional campaign. However, irrational decisions in a competitive economy usually pay a heavy penalty. The "survival of the fittest" will mean that, however the decision is made, it will be those industries which are well located that will survive and become important. A planned economy would suffer similarly from inefficient locations, although the cost of the inefficiency may be dispersed throughout the economic system. The reader is reminded, however, that here we are considering the point of view of the firm or, in the planned case, that of the project. As discussed in the introduction to this volume, regional considerations may in some cases justify higher costs from the point of view of the project in pursuing such goals as the development of a depressed region.

The majority of businessmen do not consciously go through the analysis in the form presented here. They are familiar with the operations of their industry, they know where their markets are and where their raw materials come from, and from this general knowledge can pick the likeliest spots and investigate them further. The businessman will then, if he is prudent, do a careful comparative examination



of these alternatives. He will check transport costs, frequency of schedules, labor availability, rates, quality, and organization, power costs, local taxes and prospects, available sites, climate, the housing situation, educational opportunities, availability of finance and cost of borrowing, local regulations, and any other factors of importance to his operation. This type of analysis is quite difficult, however, and most businessmen do not have within their staffs people with the training and experience to do such investigations. Consequently, businessmen are increasingly turning to location consultants to help them make a decision. These consultants do follow the logic that has been outlined, although often with differences of style.

### LIMITATIONS OF THIS ANALYSIS

We have been examining the underlying logic of the location decision from the point of view of the firm or of the project, but, of course, only in outline. We have concentrated on transport costs because these vary in a patterned way over space, and only touched on such things as labor costs and taxes, which vary in no regular way. For these, all that is necessary is to compare the savings per unit among alternatives, setting off a savings here against an extra expense there to arrive at the total localization economies. Certainly there is no theoretical difficulty here. But the theory has little to say on some important topics which remain matters of judgment rather than of scientific analysis. I shall briefly discuss some of these.

Demand has been taken for granted: we have been saying that we know where the markets are and how many units they want. This, of course, is not always a cut-and-dried matter. The demand may vary for many reasons, price often being the most important. Therefore, when a producer may sell in several markets, the delivered price of the product and the quantities sold at each of these markets may depend on the location of the plant. At the same time, the scale of operations of the plant affects the unit costs of production. Thus, it is quite possible that a plant may choose a location where transport costs are somewhat higher per unit than the minimum possible, so that it may sell its products to a city that would otherwise be outside its market area, and thus obtain a volume of production sufficient to reduce the unit production costs. And, of course, very often decisions are made in the face of competition. There may already be a firm producing the same thing at the point of minimum transport costs, so that if the new plant located there, it would be fighting for the same customers. It may be better to

---

locate at some point with higher transport costs, but where the firm can have to itself some customers that are now distant from the existing plants. Further, when a firm makes a decision it must keep in mind how its competitors are likely to react. And finally, many firms do not make just one product. A shoe factory makes many sizes and styles, and perhaps wallets and luggage. A chemical plant can turn out thousands of different combinations of products by slight changes in the processes it uses. Therefore when the firm considers alternative locations, it must also consider variations of its "product mix." The combination that may be best at one location may not be best at another.

All of these are in essence complications of the basic theory, and although the analysis may be long and difficult, fairly good answers should be possible. But there are problems of another type, involving things that are hard to quantify. Perhaps the principal one is that of *external economies*. These are the advantages or disadvantages that arise from the close proximity of the plant to other activities. For instance, a group of plants may use a machine-repair shop jointly, rather than each having its own. Among the disadvantages, higher rents and insurance rates are often mentioned. However, many of these advantages and disadvantages are very hard to measure. How much is it worth to have access to a good tax lawyer, or to be able to visit a supplier or a customer in person, or to have a first-rate printer to do a report, or to be able to receive some supplies within minutes of ordering? On the other hand, how much does it cost to fight congested traffic? It is not only hard to measure these advantages and disadvantages, it is often difficult to identify them. For instance, one large office that moved out of the city into the country was forced to return when it discovered that it could not get girls to work so far away from marriageable men. On the other hand, one research concern moved from a big city to a small town in order to keep competitors from raiding its technical personnel.

Further, and most difficult of all, are the problems of uncertainty and of time. The future is usually uncertain: tastes may change, there may be a technological revolution, tax laws and customs duties may be revised. In short, when decisions are made, it is only rarely that one can be sure of the exact results. Some interesting new work is being done in location theory to take uncertainty into account in the location decision, but so far only a small beginning has been made. In general terms, the new theory tries to estimate the costs and benefits of alternative decisions in the light of the probabilities of different things

happening. This, of course, is what businessmen themselves try to do. For instance, an American businessman will be attracted by a 10 per cent return on investment in the United States, but he will not be interested in a similar investment in an unstable foreign country unless it pays 30 per cent or more because of the danger of revolution, expropriation, severe restrictions, and so on. Thus far one of the most interesting conclusions of these investigations is that there often is no single best strategy for the businessman. He may choose to act boldly for big gains or losses, or conservatively for smaller ones, both being rational possibilities with the choice depending on the goals and attitudes of the businessman.

Uncertainty aside, the question of time itself is not sufficiently considered in the existing location theory. The decision that is best in the light of today's situation may not be best at some future time: production methods will change, and transport routes and costs, as well as the tastes, location, and resources of customers and the nature of the competition. Even when these changes can be predicted with certainty, we do not have a fully spelled-out method of making a decision that will be best over a period of time rather than at a moment in time. Such a theory would seem to be possible—in fact, not particularly difficult to develop—if we are willing to grant that all the relevant changes can be predicted with some accuracy. To combine time and uncertainty in a meaningful method of analysis will be much more difficult, but may be possible. It is clear, however, that there may be no single best answer, but that the choice would depend to a large degree on “time preference”: whether to try for early profits or for larger later ones. Both may be rational choices.

#### BIBLIOGRAPHIC NOTE

The seminal book in this field is Alfred Weber's *Theory of the Location of Industries*, first published in German in 1909, and edited and translated into English by C. J. Friedrich in 1928, (Chicago, University of Chicago Press). Edgar M. Hoover, in *The Location of Economic Activity*, (New York, McGraw-Hill Book Co., Inc., 1948), improved on many aspects of the analysis and introduced a greater measure of reality. Walter Isard in *Location and Space-Economy* (New York: The Technology Press and John Wiley & Sons, Inc., 1956), generalized and extended the theory. The works of Melvin Greenhut, such as *Microeconomics and the Space Economy* (Chicago, Scott Foresman and Co., 1963) and *Plant Location in Theory and Practice* (Chapel Hill, University of North Carolina Press, 1956) continue to explore this topic.

The reader interested in the practitioner's views may refer to L. C. Yaseen, *Plant Location* (New York, American Research Council, 1960) for the private

firm; and for government projects, to the *Manual on Economic Development Projects*, (New York, United Nations, 1958).

The example of the ice-cream vendors on the beach is based on H. Hotelling, "Stability in Competition," *Economic Journal*, Vol. 39 (March 1929), pp. 41-57. For an attempt to consider probabilities in the location decision, see W. Isard and T. A. Reiner, "Aspects of Decision-Making Theory and Regional Science," *Papers and Proceedings of the Regional Science Association*, Vol. 9 (1962).

### 3 Industrial Location and Regional Policy in Economic Development

William Alonso

#### Introduction

Most of the human race live in developing countries. There is in these countries an extraordinary range of circumstances and conditions, many stages of development, and many modes. Thus, most statements about the location of industry in these nations must be couched in the conditional. Some of the considerations that will be mentioned may seem a somewhat offensive caricature to someone from a nation already on the road to development, but I believe they will be recognizable to many from countries that are just starting this journey. The differences, however, are ones of degree, and the factors discussed appear to play a part in all countries, including the developed ones.

Further, many of the factors affecting the location of industry in developing countries are of interest to developed ones as well. In developed countries, regional planning has traditionally concerned itself with regions that are backward or depressed in economic and social terms. Many of the programs that have been tried have failed because they have not recognized the nature of the problem, namely, that these poor regions are poor because they are imperfectly integrated into the national space-economy.<sup>1</sup> The problems of regional development and the considerations affecting the location of industry in these islands of poverty are comparable to those that occur in developing countries, involving scarcities of technical and managerial inputs, antiquated social institutions and practices, uncertainties, and lack of information. Thus, the following discussion may be read either in terms of developing countries, where the modern areas are few and small and the more backward hinterland is vast, or in terms of developed countries, where the hinterland is the lesser part of the space-economy. In both cases we are dealing with relative development as a factor in location. Most of the analysis will be couched in terms of a highly simplified set of alternatives: the big city location or a hinterland location.

Given the immense variety of industries, resources, and conditions that

<sup>1</sup>The term *economic integration* refers to the geographic and sectoral mobility of factors. See B. Balassa, *The Theory of Economic Integration* (Homewood, Ill.: R. D. Irwin, 1961), and J. Friedmann, *Regional Development Policy* (Cambridge: MIT Press, 1966).

This chapter is the text of a working paper (no. 74) issued in mimeograph by the Institute of Urban and Regional Development, University of California, Berkeley, in February 1968. An earlier version of the paper was prepared in part under the auspices of the United Nations Industrial Development Organization. Most of the research and some of the writing were done under a grant from the Economic Development Administration of the U.S. Department of Commerce.

might be considered, it would be hopeless to multiply cases and examples. Rather, I have focused on this set of alternatives because it seems the most important one both for scholars trying to understand the evolving spatial structure of developing nations and for those making the location decisions. From the point of view of the businessman or the project analyst, the choice typically narrows to a consideration of two or three locations, one or two of which may be indicated because of the presence of raw materials, sources of power, a very localized and specialized market, or other such factors, but the other possible location will almost always be one of the larger cities in the country—and in many countries, there is only one large city. In most countries, moreover, the paramount issue of territorial planning concerns the possible overdevelopment of the major cities and the backwardness of the rest of the country.

This polarity is the spatial manifestation of the lack of integration of a dual economy. The relation of major cities to their hinterland is similar to that of developed nations to underdeveloped ones. The lengthy literature on “obstacles to development” which usually poses these problems at the national scale could be repeated almost unchanged at the regional scale. Thus, nations that try to develop their backward areas at the same time that they seek national economic development face a compound load of obstacles: those at the national scale and those at the regional scale.

The infant practices of regional planning and applied industrial location are often ill informed of the realities of the location process or of the true dimensions of the policy issues. Otherwise excellent studies of developing economies treat the issue of location superficially. One frequently finds, within the same discussion, contradictory statements: (a) that further concentration in the cities increases national polarization and income disparities, and (b) that this same concentration is uneconomic and leads to a slowing of national growth. It is unlikely that both of these can be simultaneously true. Or it is asserted, always without proof, that cities of 5,000 or 25,000 or 250,000, or some other population, are big enough to provide external economies. This paper lays considerable stress on these economies, but the fact of the matter is that there is no reliable knowledge of relevant urban sizes. Finally, the most common error perhaps is the bland statement that industry does or should locate so as to minimize transport costs. Except in special circumstances, this is true neither in theory nor in practice.

This paper first presents a brief and nontechnical overview of the abstract theory of industrial location. Next it deals with factors of location that are usually slighted in abstract theory, and the discussion proceeds from the point of view of the businessman or the project director. Then some questions are raised concerning the evaluation of locations from the point of view of national interest, and some considerations are suggested regarding regional planning as a factor in location. These questions and considerations only sketch some of the dimensions of the problem, since neither theory nor practice is sufficiently advanced to permit a persuasive summary.

A number of important issues will not be discussed. Among them are the priorities or linkages of agricultural and industrial development or of infrastructure and directly productive investment. These have been discussed extensively in many United Nations publications. Neither will there be explicit treatment of industrial estates or growth centers, which are as yet only a way of thinking about problems rather than a documented technique of development.

### **Brief Review of Classical Location Theory**

The formal theory of location of industry originated in Germany in the nineteenth century and remains to this day a classical theory in that it assumes economic rationality, complete information, and a static situation and ignores the complexities of the broader framework of region and nation. Thus it is poorly suited to the conditions of developing countries. Most contemporary discussion of the location of industry in developing countries does not deal with the subject as such but consists of passing references to particular aspects of location within discussions of economic development in general. Thus, many writings touch upon the overall geographic structure of a nation, upon urban and regional policies (for instance, industrial parks, poles of growth, urban decentralization), or upon transportation planning.

This paper will try to draw together many of these scattered insights into the forces acting upon industrial location in developing countries. Although classical location theory is ill suited to developing nations, it is useful to have a brief, nontechnical review of it at this point to start the discussion.<sup>2</sup> Further, this exercise is of interest because its deceptive straightforwardness molds the thinking of many practitioners and scholars of development, principally through the half-truth that the rational location for industry is that which minimizes transport costs.

The most classic problem in the literature is that of choosing a location for a factory. It is assumed that the sources of materials are known, as well as the location and size of markets, the necessary quantities of the several materials per unit of product, and the relevant transport rates. When the problem is so stated, everything but transport costs is held constant; consequently, the best location is, quite obviously, that which minimizes transport costs. We shall see later on that other things do not stay constant, and that this objective must therefore be heavily qualified. For the moment, however, we will accept the classic formulation.

Consider the costs of assembling materials and distributing the product of a manufacturing plant at each possible location. These costs will vary from one location to another, and taken together they will describe over the national territory a fairly smooth surface of transport costs, shaped like a bowl, with some pits in it. The lowest point of that surface must fall within the

<sup>2</sup>A more formal and complete summary is available in W. Alonso, "Location Theory," in *Regional Development and Planning*, ed. J. Friedmann and W. Alonso (Cambridge: MIT Press, 1964), reprinted as chapter 2 in this volume.

polygon formed by drawing lines joining the locations of markets and materials at one of the vertices of the polygon. The point of least transport costs may be the lowest point on that surface, but the surface is likely to have pits at each material or market location. This is because locating at such points will avoid the terminal costs for a material of loading or unloading, taking of insurance, arranging for shipment, etc., by locating at its source—or for the product by locating at its market. These pits may be lower than the bottom of the bowl and may thus represent locations of lower transport costs.

There is another type of location where savings of this type are possible: namely, ports and other points where there is transshipment or break-of-bulk. It is often advantageous to process materials at the time they are taken, say, out of trucks to be loaded onto a ship. This, of course, is one of the principal reasons for the development of port locations as industrial centers. In the terms we have been using, port locations often represent deep pits in the surface of transport costs.

Classical theory recognizes that transport costs are not usually the only consideration. Certain locations may have lower wage rates, preferential tax treatment, or some other advantage that lowers the costs of manufacturing from the businessman's point of view. The savings per unit of product may be subtracted from the height of the transport surface at that point, causing another pit which should be compared to the others.

In short, this simplest formulation of the theory amounts to no more than a systematic consideration of all possible locations, calculating transport and production costs at each location, and choosing the lowest. It may be noted that it is more of an engineering than a true economic model, at least in this simple form. The shallowness of the conclusion that the best location is that which minimizes transport costs (or, in slightly more sophisticated formulations, transport plus production costs) may be illustrated by analogy. Consider a firm for which we know all revenues and costs except those of capital. Then, the optimal quantity of capital would be the minimum. Yet it would be obviously ridiculous to say in consequence that the optimal solution would minimize the amount of capital investment if we allowed the other factors to vary. Similarly, the commonly accepted role of minimizing transport costs does not take into account economic considerations such as factor substitution, the elasticity of demand, economies of scale, or the consequences of alternative pricing policies. Full consideration of these would take us to a complex mathematical analysis,<sup>3</sup> and only a suggestion of these effects is possible here.

Diverse locations will result in different delivered prices for materials, and many manufacturing processes will therefore adjust their factor proportions. For instance, a steel-making plant at a location distant from a source of coal would find coal relatively expensive, by reason of its additional transport

<sup>3</sup>The interested reader is referred to W. Alonso, "A Reformulation of Classical Location Theory and Its Relation to Rent Theory," *Papers of the Regional Science Association* 19 (1967).



costs, and would probably choose a production process that used relatively less coal and more of something else. Thus, the classical theory studies substitutes only among transport inputs, whereas in reality there often exist possibilities of substitution between transport and other inputs. The objective of minimizing transport costs must be broadened to the minimizing of total costs, and this only if we take total revenues as fixed.

But total revenues are not fixed if the quantity to be sold varies with the price. If the product is sold at different prices in different markets because of differences in delivery costs or other reasons, the profitability of the enterprise will generally be greater by locating near those markets that are sensitive to price variations. Thus, a large market that has an unvarying rate of sales will exert less attractive force than a smaller one in which slight reductions in delivered price result in large variations in the quantity sold. The quantity sold at each market may thus vary from location to location of the plant and also with pricing policy adopted. This point is very important, for it is profit, not the minimizing of costs, that is a truer objective of most economic activities. The reduction of transport costs is only a partial objective, and the firm will normally be willing to spend more in order to make more. Otherwise, in the extreme case, transport costs would be minimized by not producing at all.

It must be recognized, however, that the classic theory's problem of production to a predetermined quantity and price may be more suited to conditions common in developing countries: scarcity, imperfect articulation of the pricing system, and socialized production. It may be argued that in those cases, the problem is to engineer production rather than to identify demand, in contrast to the free-market situation of affluent economies. But such an argument and the objections to it lead to rather difficult and as yet poorly defined areas of economics. A simpler and, to my mind, more forceful argument for a location that minimizes transport costs is that the analysis is simpler, less demanding of limited technical manpower, less apt to attract the push and pull of political and other interests, and therefore likely to be less time-consuming, so that the project may get under way more quickly at a satisfactory location even if that location is not the best in terms of profitability, contribution to national product, or some other more fundamental objective. It may be noted that different objectives will sometimes result in different locations. (This will be discussed in the next two sections.)

Under the terms of the theory that has been presented, the location of the firm may be viewed as responding to pulls from markets or materials. In the simplest case, where all quantities are fixed and only transport costs vary, the pull of a source of material is equal to the weight of that material per unit of product times its transport rate. The pull of a market is equal to the fraction of total production sold there times the weight of the product times the transport rate. If we allow substitution among production factors, the pull of a material is weighted by the sum of cross-elasticities between this factor and the others. Under certain methods of pricing, the pull of a market is weighted

by the price elasticity of the demand at that market. But, such complexities aside, we may imagine these various pulls operating on the location: the optimal location will be at that point where these forces balance each other and equilibrium is achieved.

This simple, physical analogue of the forces of location is quite instructive. The force that pulls toward a location is called the *ideal weight*.<sup>4</sup> Where one ideal weight is large in relation to others, it pulls the location of the plant toward it, and the industry is said to be oriented toward it. Thus, processes that lose weight (such as the reduction of ores) result in a *material-orientation*. On the other hand, processes that increase in ideal weight tend to be market-oriented. The increase in ideal weight may be, as in the case of beer and ink, due to the addition of a universally available ingredient (water), or, as in the case of boxes and automobiles, due to the increase in bulk which increases the freight per mile, or, as in the case of newspapers and baked goods, due to the high perishability of the product which requires rapid and expensive transportation.

Even these rudimentary considerations begin to shed light as to whether industry will be attracted to resources. Thus, iron ore has typically a low ideal weight in relation to coal and other material factors that go into iron and steel; the higher the grade of ore, the less likely it is to attract industry since it is easier to move. Oil has an ideal weight nearly equal to that of its first-stage products; therefore, petrochemical industry is free to locate either at the source, or at the market, or at some intermediate location that is convenient because of transportation, availability of funds, political safety, or any other pertinent reason. One may generalize that most physical resources attract little industry. Their first elaboration does not greatly add to their value and is seldom labor-intensive, and later transformations are likely to be oriented to other locational pulls. On the other hand, coal is weight-losing in most of its uses, and electricity is expensive to move over long distances. Hence their ideal weights are high and they may attract industry.

In comparison with industries in developed countries, most industries in developing countries are simple in that relatively few raw materials are required. Thus, the steel industry will devote itself to the basic steel types rather than to more specialized alloys. At the same time, the markets for the products will be concentrated in a few locations, typically the major cities, or, in the case of foreign markets, the ports of export will serve as the terminals within the national territory, equivalent in a sense to a market. A complex process that uses materials derived from many primary sources and secondary producers and that produces a wide line of goods which sell in a multiplicity of markets is subject to many locational pulls and often finds itself relatively free as to where it will locate since an increase in distance to one market or material is likely to be compensated by a smaller distance to another. In other

<sup>4</sup>Actually, ideal weight is the name of the magnitude of the force (excluding direction) in the simple case. No general name exists for the more complex directed force.

words, the bowl of the transport surface for processes using many materials and yielding several products is likely to be rather flat. By contrast, where the locations of markets and materials are few, the transport surface is likely to have a very distinct low point, and therefore it is likely that there will be a much more definite optimal location in the case of developing countries. This tendency is strengthened by the fact that the ratio of the value of materials to the value of final product is typically far higher in developing countries. That is to say, the value added in the production process being lower, the location of the plant is more likely to be determined by differences in transport costs. We shall see, however, that a host of other factors in developing countries counter the importance of transportation.

#### **Location of Industry in Developing Countries from the Point of View of the Firm or the Project**

This section deals primarily with the firm, although almost all of the criteria presented will apply also to government industrial projects if they are viewed as projects only, rather than as part of a general policy of national spatial development. (Subsequent sections deal with changes in some of these factors when viewed from considerations of national policy.) These criteria apply particularly when a firm is both owned and operated by an individual or a family, as often happens in developing countries. Some of the factors discussed, such as the personal space preferences of managers and technicians operate less strongly when the firm is operated by people other than the owners, and it will matter less whether the owner is the nation or some large corporation or a group of shareholders.

Further, we will deal with firms that may be said to be "located"; firms for which the location decision is made consciously. Little will be said about what may be called "spontaneous" industry; that is, industry which owes its location to the happenstance of the place of residence of the person who started it and to his ability to see or to create his opportunities. Although a great many countries, both developed and developing, have shown the phenomenon of regions or cities in which new industries spring up from local sources, little seems to be known about the reasons for this fertility. Much of the fundamental work belongs to an earlier generation of economists. Among present-day students interested in this phenomenon, Everett Hagen holds an important place.<sup>5</sup> In general, the conditions that bring about this ferment appear to involve a certain tension or restlessness, a dissatisfaction with things as they are or some threatened changes, which cause certain groups or individuals to look for opportunities and ways to introduce change. A racial or religious minority, an uprooted section of the farm population, a threatened landed aristocracy have all, on occasion, played the entrepreneurial role. Yet generalizations as to the spatial incidence of this

<sup>5</sup>See, for instance, E. Hagen, *Planning Economic Development* (Homewood, Ill.: R. D. Irwin, 1963).

spontaneous industrialization are most difficult, and no way is known to plan for it. It may be noted, however, that where such spontaneous industrialization is successful, it is likely to lead in time to the professionalization of management and to the separation of management from ownership. As the scale of operations increases, the location decision will become more explicit, especially for branch plants, and what starts as a "spontaneous" industry may become a located one.

We are dealing here with differences of the location decision between the classical theory and the conditions of developing space-economies. It will be seen that the principal differences may be termed differences in information, which involve uncertainty, knowledge of opportunities, technical and managerial capacity and supply, and other factors. Most broadly, it can also involve the movement of goods as well as ideas. The classical location theory, by assuming perfect information and predictability and unlimited managerial and technical energy, ignores the strongest factors which operate in a context of development. Similarly, classical theory is for the most part static, in effect assuming that things happen in an instant. We shall see that in developing countries time plays a crucial role.

We shall begin by examining the effects of the spatial structure of developing countries upon transport costs, and then generalize the discussion to include some vital but often neglected factors.

#### **The Transport Network and Concentration of Markets**

The transport networks of developing countries reflect their economic history of being exporters of primary products. The resulting transport networks resemble drainage systems converging toward the coastal ports which became the major cities. These served as transshipment points not only for exports but also for the distribution of imported finished goods. The transport network, therefore, is typically shaped like a fan or a tree.<sup>6</sup> In developed countries, by contrast, the shape of transport networks is a fuller lattice which, by offering more nodes or points of confluence, permits a more even distribution of locations. A glance at world maps of population concentration shows a striking pattern of coastal cities ringing the developing continents, while the hearts of the continents remain virtually empty. In some cases, a navigable river permits some development of a city further inland, but this is really a fortunate penetration by the sea into the continent. If instead of concentration of population, the concentrations of industry, income, education, or other forms of economic development are examined, this ring effect is even stronger.

The consequence of coastal concentrations and fan-shaped movement systems is that the point most accessible to the country as a whole is not its

<sup>6</sup>See P. Haggett, *Locational Analysis in Human Geography* (New York: St. Martin's Press, 1966), in particular for his discussion of the works of Taaffe and Kansky. The most detailed and thorough documentation is in B. J. L. Berry, *Essays on Commodity Flows and the Spatial Structure of the Indian Economy*, Department of Geography Research Paper no. 111 (Chicago: University of Chicago, 1966).

geographic center, but a point on its edge, which is the functional center of the country in terms of economic distances. It may be noted that the physical eccentricity of these functional centers results in longer average distances among points in a country of a given size; on the other hand, the extreme concentrations of economic activity that characterize these countries reduce the distances which are effectively traveled.

One consequence of this strong directional veining of the national territory is that in the preliminary analysis of location, distances must be considered in terms of specific routes; in developed countries, they may be taken more abstractly. In a developed country—in Europe and North America—one can be relatively certain that the effective road distance between any two points on the map will not exceed the air distance by more than 20 percent. In developing countries, two relatively neighboring locations may lack a connecting road and be joined only at some distant major city. The technical elaboration of this concept is beyond this paper, but the point being made is that full analysis of location in developing countries will involve either more difficult geometric transformations of physical space into economic space than is the case in developed countries, or that more specific and detailed (and therefore more demanding) analysis will be needed in the preliminary stages of planning.<sup>7</sup>

The large coastal cities are thus the most central locations for distribution of goods to the population at large, and this is a factor of prime importance for industries producing consumer goods if the scale of their market is nationwide, as is likely due to the limited demand.<sup>8</sup>

The large coastal cities or principal nodes will also be most central for industries that depend on one material assembled from dispersed sources, or that combine diverse materials from diverse sources. This centrality is reinforced by the port function, of course, when some of the materials, components, or equipment are imported, or some of the product exported. Further, the concentration of demand at that point is often preponderant, both because income is always more concentrated geographically than population and because manufactured products are typically income-elastic, so

<sup>7</sup>W. Warntz has a most interesting way of looking at this. Human modes of interaction respond to the difficulties of overcoming distance (Haig's "friction of space"), and any one mode of interaction, such as bank transfers or the travel of people, forms a hypersurface that may be visualized as crinkled cellophane over the geographic surface of the territory. The distance between any two points in a system or mode of interaction is the geodesic (least distance path) on that hypersurface. Economic development and technological innovation tend to cause these hypersurfaces and geodesics to conform increasingly to conventional maps and geographic distance. Glimmers of these concepts, which Warntz has discussed with me, appear in his "Global Science and the Tyranny of Space," *Papers of the Regional Science Association* 19 (1967). It will be apparent that this conceptualization, while more abstract, coincides with Hirschman's cycle discussed below, with the concepts presented in W. Alonso, "Urban and Regional Imbalances in Economic Development," *Economic Development and Cultural Change* (forthcoming), and with Friedmann's and Belassa's concept of integration of the space economy.

<sup>8</sup>On the other hand, consumer industries that can operate effectively at small scale can and should be distributed more widely over the national territory.

that demand is more concentrated than income. Similarly, industrial demand for producer goods will typically be concentrated at these points.

We can see, therefore, that transport-oriented industries will often find their most advantageous location at these coastal nodes. It is a contention of this paper, however, that the minimizing of transport inputs is a lesser factor in the location of industry in developing countries than in developed ones. Rather than transport orientation, there is an attraction to production economies of a subtle type most often associated with externalities, which serve as powerful magnets, drawing location to the big cities. This is not to say, of course, that many industries do not determine their location by the classical location factors associated with transport costs. Most typical of these is first-processing of materials, which is a weight-reducing process and therefore binds the location to the source of the material. Sargent Florence cites among these the earlier processes of metal refining, cotton ginning, bacon curing, seed crushing, sugar refining, sawmilling, rice milling, and cement and brick making.<sup>9</sup>

#### **Distance as Time**

Often, time rather than cost is the prime consideration in distance. This is well known in the case of those industries that deal with perishable commodities. In developing countries, however, other factors make time important for many more industries.

The time consumed in simple shipments can be surprisingly long. It is not uncommon for it to take weeks to move commodities or materials over relatively short distances, with days wasted in confusion, breakdowns, layovers, unavailability of carriers, labyrinthine paper work and authorizations, occasionally washed-out roads or tracks, and a thousand other contingencies that defy the imagination. Sometimes, after weeks of waiting for a shipment, it is discovered that the order was misplaced and the shipment has not left. Or the shipment itself may be lost, in the sense that it has arrived at some place and no one has been notified of the fact. The probability of these occurrences will vary greatly from country to country, but it may be generalized that in developing countries shipments take longer on the average and that there is greater uncertainty as to the time of arrival.

Under these conditions, the time consumed in transportation is likely to be an important consideration for many industries. At distant locations, it will be necessary to have substantially greater inventories of materials to insure continuity of production and also necessary to have large inventories of the product at the markets to insure prompt and predictable deliveries. Even the inventory of spare parts for the production machinery will have to be greater. This considerable item has been estimated at about 20 percent of total machinery and equipment investment in developing countries.<sup>10</sup>

These delays, then, require considerably larger investments of capital for

<sup>9</sup>P. S. Florence, *Economics and Sociology of Industry* (London: C. A. Watts, 1964), p. 151.

<sup>10</sup>M. Bryce, *Industrial Development* (New York: McGraw-Hill, 1960), p. 125.

inventories at distant locations, to cover both the longer lead times involved in the length of normal trips and the possibility of unforeseen delays. The additional capital costs, given the high interest rates that prevail in these countries, can easily exceed savings in transport costs as such. The first type of cost, involving lead times, is unavoidable, especially in the absence of complex and sophisticated inventory techniques. The second type, arising from uncertainty, is avoidable only at the probable cost of idle production lines if materials or spare parts fail, or of dissatisfied customers and unfilled orders if the product cannot be delivered.

#### **Personal Space Preferences of Managers and Technicians**

Personal preference, perhaps because it is so human, seldom receives adequate recognition, yet it may be in some ways the most powerful one in operation. Managers and technicians typically belong to certain social classes, such as the aristocracy or an emerging upper-middle class, have usually had a relatively good education, and expect a way of life which can be had only in the principal cities. Varied restaurants, movies, clubs, interesting friends, well-dressed people, fashionable shops, bookstores, television, a sense of being where things are happening: these are felt to be necessary by the majority of these men and the overwhelming majority of their wives. A location away from the big city is often regarded as exile.<sup>11</sup> Family links, which are of great importance in many developing societies, will often be another anchor to the big city, unless the man or his wife happens to have family at the alternative location.

Aside from these ties to the big city (which this man, even if born elsewhere, will regard as his home since he probably spent many years there as a student), there are likely to be serious problems of personal and family adjustment at the new location. These may include ethnic or religious differences, a different dialect, a local aristocracy to which he will be an outsider, different customs, different food, etc. It may be hazarded that the family's adjustment will be made more difficult by their feeling superior to their new environment; they are likely to view it as not merely different, but also as provincial and old-fashioned.

To get managers and technicians away from the big cities, very often substantial inducements must be offered, including higher salaries, company-provided houses and automobiles, lengthy vacations, and paid trips. Even at that, it seems likely that first-rate men will often not be obtainable, for they have ample opportunities in the big cities. In developing countries, where the newness of situations and the lack of an institutional matrix call for the highest order of resourcefulness, energy, and inventiveness in these positions, a second-rate man can be far more costly in lost efficiency and missed opportunities than any differential in transport costs. More will be said about the special calls on management and technical staff in emerging industries. It

<sup>11</sup>This phenomenon is not limited to developing countries. A recent survey of this issue by the *Wall Street Journal* (February 23, 1966) found it a considerable factor in determining locations for American industry.

seems clear, however, that such personal preferences will play a greater role in firms that are owned and managed by an individual or a family than in plants that are part of enterprises with a more abstract ownership, as in the case of government projects or publicly held private corporations. In these, management is professionalized into a career, and the decision about where to locate a plant is usually made by people who will not go there to live.<sup>12</sup>

#### **Spatial Distribution of Information and the Cost of Time in the Location Decision**

A common feature of developing countries is that reliable information about possible hinterland locations is often difficult or impossible to obtain. Not only may it be hard to find out about transport facilities and schedules and the availability of supporting services and facilities, but it may be difficult to learn about possibly important features such as the properties and availability of water, local woods, clays, and other materials. Local variations may require redesign of equipment and processes, often involving scientific investigation which may be lengthy and not necessarily successful. There is frequently the possibility that something that has been taken for granted will not be available and that the firm will be put to the expense and waste of energy of going into these activities to supply itself. The possible list of things that may be missing is too varied to do more than illustrate it. The firm may be forced to become its own trucking and shipping firm, a warehouser of office supplies and repairman of office equipment, a real-estate manager and general store for some of its employees, its own jobber for certain materials and components, etc. These, of course, are the opposite of external economies and may be termed diseconomies of forced internalization. Several countries that have essayed industrial decentralization policies have discovered that large firms can more successfully adapt to the conditions of small cities and distant locations. Their size affords them economies of scale for internalizing these externalities.<sup>13</sup> We shall discuss below other factors, such as the predictability of demand and stability of production processes, which affect a firm's ability to adapt to a hinterland location.

A particular point deserves discussion here, although it might be reserved for the final section. Aside from classical factors such as lower transport costs or localized advantages, one of the principal reasons for advocating the location of industry in the hinterland is the national goal of equity, which tries to minimize differences in income and employment among regions. Experience in both developed and developing countries, however, indicates that

<sup>12</sup>See A. Lauterbach, "Managerial Attitudes and Economic Development," *Kyklos* 15, no. 2 (1962). While making similar although nonlocational observations, Lauterbach observes that industrial enterprises of powerful families in Latin America are often associated with their agricultural background and holdings. Thus, they tend to develop canneries, sugar refineries, coffee roasting meat freezing and plants, and wool and cotton mills. Both the economics of location and the space preferences of some of these families may often favor hinterland locations.

<sup>13</sup>See External Economies, below.



frequently much of the labor for a new enterprise must also be brought into a laggard region because the local population lacks necessary skills. In these cases, except at the most superficial statistical level, the goal of interregional equity is not served, for the newly employed are not, except post facto, the residents of the region. The original population may receive no benefit except for multiplier effects, which will be small because of sparse intraregional sectoral linkages and which may, in fact, reduce local welfare by resulting in sharp sectoral imbalances within a region with low productive adaptability. In other words, the resulting increases in demand for food, construction materials, and the like may raise their price to the local population.

The most common advantages of distant locations may lie in lower transport costs, lower rents, lower wages, a labor force that is less set in its ways, and fewer costs of congestion. Contrasted to these are such disadvantages as the risk of unforeseen costs arising from local conditions and the costs of internalizing functions which normally would be external and for which the firm may lack efficient scale. These costs are largely unknown at the time of making the location decision, and they are evaluated probabilistically in any rationally made decision. Since profits are, ultimately, the difference between costs and revenues, greater uncertainty as to the upper limit of costs will require a larger difference for safety, especially since revenue is usually more predictable from the price of competitive products. In consequence, the known advantages of the distant location will have to be very large to offset the probable, but unknown, disadvantages. In other words, uncertainty or possible error is not symmetrical about the most probable estimate of costs, but rather is strongly skewed in the positive direction. The probable profits, therefore, are strongly skewed in the negative direction, and any sort of strategy for minimizing losses or maximizing profits will tend against situations with higher uncertainty. Thus, the neglect by businessmen of peripheral locations is based on probabilistic rationality.<sup>14</sup>

The very high output-capital ratio and rates of interest which normally prevail in developing countries provide yet another force toward the known environment of the large city. Even if it were possible to get the information necessary to remove all uncertainty about a distant site, it would probably take months or even years to do so. Thus, to explore this possibility involves a delay in getting started. When returns to capital and entrepreneurship are very high, delay may be quite costly in terms of foregone production. Thus, the choice will often be reasonably made in favor of the known opportunities in developed locations *now*, even when it is recognized that there exists a strong possibility that another location would be better if there were time to look into it. Vigor and decision can be better than careful thought if a satisfactory alternative is available and finding the optimum causes delays.

Returning to the matter of labor costs in the perspective of time, it must

<sup>14</sup>This estimate of the rationality of favoring location in the major centers differs from A. Hirschman's in *The Strategy of Economic Development* (New Haven: Yale University Press, 1958), p. 185. See below.

be noted that only in a relative long run is cheaper but less-skilled labor usually an advantage. In the first years, as labor is being trained, one may expect it to be inefficient not only because of lack of specific skills but also because of lack of industrialized attitudes. Although subtle in their fundamental forms, attitudes may manifest themselves in such expensive specifics as high rates of absenteeism, tardiness, turnover, spoilage, and poor maintenance of equipment.<sup>15</sup> Given the very short financial planning horizons that are typical of developing countries, a delay of a very few years in reaping the benefits of lower wages will destroy the attraction of that factor. On the other hand, as will be discussed later, public policy, which often operates with longer-range planning horizons and should be interested in things other than the balance sheet of a project, may regard the creation of modern attitudes in the population to be as important as the physical production of goods.

#### External Economies

External economies, which are among the most important determinants of location in developing countries, are the most difficult to treat. There exists no comprehensive systematic categorization of them nor sharp tools for their analysis or the measurement of their effect. There is an ample descriptive literature, but it is limited primarily to the developed countries. We shall see that certain aspects are of particular importance in developing countries. Although they are hard to measure, they must not be underestimated. There is considerable evidence that in many cases they can more than compensate for higher costs of transport, labor, or other factors of production from the point of view of the firm or the project. From the point of view of national regional development policy, an understanding of the elements of external economies probably contains the answer to the crucial questions of whether the principal cities are too big, of how big secondary growth centers must be to enjoy self-sustaining growth, and of what types of industry are proper subjects for a policy of decentralization and at what stage of their evolution. These policy issues will be mentioned in greater detail in a later section.

We may begin by following a classic statement of external economies. Hoover points to three elements in external economies: multiples, massing of reserves, and bulk transactions.<sup>16</sup> These are amplifications of the same concepts which had been applied to the efficient size of firms. In the case of the firm, the concept of multiples implies that, since machinery is designed to operate at certain capacities, the minimum efficient size for a firm is that which insures that no piece of equipment is underutilized. By way of illustration, imagine a three-step process: if one machine at step C accommodates the output of two machines of step B, and one machine at step B

<sup>15</sup>Although these phenomena are commonplace and familiar in developing countries, it is interesting that some of the most precise documentation is for a developed one. See W. F. Luttrell, *Factory Location and Industrial Movement* (London: National Institute of Economic and Social Research, 1962).

<sup>16</sup>E. M. Hoover, *The Location of Economic Activity* (New York: McGraw-Hill, 1948), pp. 120-121. Hoover refers this categorization to P.S. Florence, *The Logic of Industrial Organization* (London: Kegan, Paul, Trench, Trübner, 1933), who applied it to the firm.

accommodates the output of three machines of step A, the minimum efficient size of a plant would consist of six machines of step A, two machines of step B, and one machine of step C.

The same logic applies to a city. A certain minimum size is required for the efficient use of an airport, a transport system, or other forms of infrastructure. Smaller sizes will mean overcapacity in performing one of these functions, and therefore greater unit cost. At the same time, and perhaps more importantly, certain sizes must be reached to justify the coming into existence of specialized supporting services, such as specialized shippers and jobbers, financial agencies, trade publications, repair services, specialized printing, consulting services, equipment leasing, advanced education facilities, laboratories, etc. It should be noted that the principle of multiples, in its full sense, may result in higher productivity not only through the avoidance of idle capacity but also through permitting firms to specialize more narrowly and thus increase productivity. A firm could print its catalogues at a printer who is especially set up and experienced in that type of work; it could seek advice from a lawyer who is particularly knowledgeable on the issues at hand; it could get its machines repaired by men who know them well and have the proper tools; and so forth.

The principle of massing of reserves is an actuarial principle. Imagine, for instance, that there are five firms, each of which plans to use ten units of a material, but each of which, because of the possibility of spoilage or unexpected orders, considers that there is an even chance that it will need an additional two units. If these firms are separated in space, each must have twelve units of material for that level of safety, and the five will need a total of sixty units. This constitutes a reserve of ten units, or 20 percent. On the other hand, if all five are close together and use a common supplier or have an understanding to assist each other, only four and one-half units, or 9 percent of the material will be needed as a reserve for the same level of safety.<sup>17</sup> The same logic applies to the advantages of larger pools of skilled labor, larger markets for the product, and larger or more numerous suppliers of the factors of production, repair, and other services. The application of this principle to labor is of particular importance, for it means that, aside from economy-wide cyclical fluctuations, a low rate of unemployment in a big city provides its industries a greater margin of safety in labor reserves than does a higher rate of unemployment in smaller centers. Conversely, layoffs by one firm in a small city are far less likely to be absorbed by other firms. Therefore, nations that seek growth along with full employment should recognize that a policy of decentralization is likely to carry with it a higher rate of idle labor.

Whereas the principle of multiples refers to the determinate advantages of particular combinations, the principle of massing of reserves deals with the probabilistic advantages of sheer size. It is based on the fact that uncertainty,

<sup>17</sup>This is based on the independence of each firm's need for extra material.

when spread over large numbers, is more predictable and therefore less risky. Looked at another way, large numbers permit much greater flexibility and adaptability to changing and often unforeseeable circumstances because more opportunities are open. For instance, if a firm changes its production process and thereby finds itself with a new waste product, it is far likelier in a large city that there will be some other firm which can make use of it than in a smaller city, thus transforming the waste product into a by-product. Similarly, if the new process calls for a new input, it is far more likely to be obtainable in a large city than in a smaller one.

The third element in this formulation is the principle of bulk transactions. In the case of the firm, this is based on the fact that buying or shipping in larger quantities usually results in lower unit prices.<sup>18</sup> In the case of the city, the large scale of operations may result in lower rates and more frequent and convenient service for large-scale transfer and terminal facilities. The same may apply to other public and commercial services, principally because fixed costs may decline per unit and because more efficient processes can be used at larger scales. It is frequently suggested, however, that many functions in large cities operate under diminishing returns, principally because of congestion, either internal to the function or among functions. This point has not been settled by empirical research, and it constitutes the principal bone of contention between those who argue for big cities and those who argue against them. I shall have more to say on this later.

The principle of bulk transactions has a less equivocal advantage, already encountered under the principle of multiples. A very large scale of operations permits specialization and increased productivity: the very large number of events and transactions make it possible to handle a wide variety of needs in a standardized and therefore more efficient and less costly manner. We have reference again to specialized shippers, repairmen, professionals, subcontractors, etc. This does not contradict my argument below that the large city is particularly hospitable to the unstandardized producer. The reason for this is that the unstandardized producer has greater need for supporting services and associated activities which are found in the big cities; further, in large cities they will be more specialized and standardized and therefore easier to use.

The subtle but enormous importance of face-to-face relations is not sufficiently recognized in the preceding formulation of external economies. Yet these personal contacts are crucial in developing countries for many reasons. At the most obvious level, developing countries have poorer communications systems. Telephones are scarcer and less dependable, mail is slower and may fail to arrive, air connections are rarer. Consequently, distance between people is a more formidable barrier to communication and this leads to the spatial concentration of those who must do business together.

<sup>18</sup>This is equivalent to T. Scitovsky's "pecuniary external economies" in his classic "Two Concepts of External Economies," *Journal of Political Economy*, April 1954.

Social forces, however, may exert an even stronger force toward spatial concentration. In developing countries most messages are less impersonal and standardized than in developed countries where much information is transmitted by trade journals, catalogues, government publications, and other means. At the same time, in developed countries, contracts, terms of finance and details of payment and delivery, specification of the product, and a thousand other forms of contact tend to follow established forms. In developing countries, communications depend far more on word of mouth, and procedures are far less standardized and call for discussion and negotiation. These circumstances require people come together. Further, the rituals of social contact when two-way communication is necessary are more elaborate. In many countries considerable time must be devoted to expressions of mutual respect, and no haste to get to the substance of the meeting may be shown. Even the substance of the conversation must be treated with a subtlety and indirection which often baffle those used to procedures in developed countries. There are ways of saying yes that mean no, price and time estimates that mean something else, gentle probings and subtle hints. An order blank, a long-distance telephone call, a cable, or a business letter are instruments too blunt and insensitive for these complexities.

Mechanical problems of communication and the customary modes of contact are not the only forces of this type that draw firms toward the city. The social and institutional matrix of communication and information also exerts pulls. Personal relations enter more frequently and more pervasively into human contacts of all types, including business relations. In developed countries, contacts are more segmented into well-defined roles which are limited to the issues at hand.<sup>19</sup> In developing countries it is often more important to cultivate sources, to establish a commonality of attitudes or class positions. Information on new developments becomes available at irregular intervals, and channels must be kept open to be certain of being informed. Since the information is unpredictable, often taking the form of rumors of confidential reportage, it is important to be in contact with many people and to see them even when no particular business is at hand. Conditions in developing countries change frequently and unpredictably: regulations are changed, ministries are reorganized, the relative costs and availability of production factors and financing shift suddenly. It is therefore often crucial to be at the centers of influence, decision, and information. The complex bureaucratic maze of permits and licenses requires that papers be personally shepherd from one official's desk to another. This is not only because rules and procedures are often confused or even contradictory, or because many an overbureaucratized machine needs to be lubricated by flattery, persistence, and great resources of persuasion, including bribery, but also because the low standards of professional competence of many

<sup>19</sup>This, of course, is a matter of degree. For instance, even in developed countries, salesmanship is very often pursued by warming personal relations in the sharing of meals or leisure.

government officials below the very top levels make it likely that papers will get lost on some desk or dusty file in the depths of a ministry.

I have argued that the fluidity of the situation, the lack of standardization of procedures, and the personalization of contacts pull forcefully toward the city. This pull is magnified by the scarcity of entrepreneurs and managers. In most firms, one or a few individuals have all the responsibilities of management, including general management, marketing, finance, purchasing, production, and long- and short-range planning. For the reasons discussed, the management functions are pulled toward the city, and they bring the production functions with them. The advantages of flexibility, adaptability, and the conservation and efficient use of the scarce time of the management are likely to wash out any advantage of a few percentage points in transportation, labor, or some other cost at an alternative location.<sup>20</sup>

It is enlightening to compare these attractions to industry in developing countries with the history of new industries in developed countries.<sup>21</sup> These new industries typically spend their early years in the large cities for reasons that are similar to those which I have been discussing: neither production nor demand is standardized and situations change rapidly; ownership, management, and technical functions are combined in one or a few individuals; the instability of demand and frequent changes in technology or the cost or availability of production factors put a premium on contacts and adaptability; investment per worker is low since larger commitments become wise only when the situation and the processes are stable and predictable enough to justify with safety extended runs of identical items; in the early stages of the industry, it is vital to have access to common services and inventories outside the firm, to sources of finance, and to complementary producers. In short, these new industries are very open systems, oriented to external economies.

With time, demand becomes more predictable, production processes and product lines are stabilized, the various functions of management become professionalized, specialized, and distinct, and this greater stability permits routinizing and internalizing of some of the services and steps of production which earlier were done outside the firm. In short, the firm becomes more of a self-contained system, internalizing many of the functions for which it had depended on externalities. At this time, the industry often moves out of the large city toward a location that may offer cost advantages in transportation, labor, or some other factor. Even the activity of innovation often becomes institutionalized in research sections staffed by professionals who work systematically toward defined objectives, and this activity may be quite free

<sup>20</sup>Related ideas are discussed in H. Leibenstein, "Allocative Efficiency vs. 'X-Efficiency,'" *American Economic Review*, June 1966.

<sup>21</sup>See R. Vernon, *Metropolis 1985* (New York: Doubleday Anchor, 1963), and R. M. Lichtenberg, *One Tenth of a Nation* (Cambridge: Harvard University Press, 1960). Valuable documentation and related ideas are found in A. Pred, *The Spatial Dynamics of U.S. Urban Industrial Growth, 1800-1914* (Cambridge: MIT Press, 1966).

from locational constraints, if not from locational preferences. It is also frequent to see the formation of branch plants separated from the head office, since the professionalization (depersonalization) of roles and the routinization of communications permit control of production, deliveries, etc., by standard messages.<sup>22</sup>

In a newly industrializing country most industries will be "new industries." Even those that in developed countries are well established and standardized may, in the developing country, be expected to undergo an extended period of adaptation as management, technicians, and skilled workers learn their business, as production methods are adapted to local conditions, and as demand defines itself. The most commonly noted of these adaptations arises from the differences in relative costs of factor prices in developing countries. Typically labor is less expensive and capital more expensive, and the technology may have to undergo a difficult adaptation to a more labor-intensive process. During this period of adaptation, the city will usually be the most hospitable seedbed. As the industry matures, it will exhibit a greater tendency to seek other locations, although it may be detained by the factors already noted of poor communications, traditions of direct personal contact, etc.

Countries starting on the road toward industrialization will, by definition have a larger proportion of "new industry" that will contribute to an apparent overurbanization of excessive concentration. With development and a larger base of established industry, the new industry that seeks the externalities of the city will constitute a smaller proportion of all industry, and the location of industry as a whole will become more decentralized. At the same time, as the nation advances toward economic development, one may expect changes that will facilitate such relative decentralization. Thus, the quality of infrastructure and of information about local conditions will become more even across the national territory, as will the adaptability of the population to modern conditions of production by reason of increasing literacy and habits more suited to the required roles.<sup>23</sup> A sectoral shift may be expected in the structure of the national economy, with a typical relative shift from the dominance of raw materials for export and easily produced consumption goods to a more significant level of processing materials within the national territory and, ultimately, to greater production of intermediate and capital goods. Insofar as the productive structure deemphasizes imports of intermediate goods and consumer goods and as exports of materials decrease in relative importance, the great cities will lose some of their attractive powers as ports. On the other hand, the tendency toward relative decentralization

<sup>22</sup>It may be hazarded that most countries, both developed and developing, that have tried to decentralize industry have had more success with distant locations for large, integrated complexes that are largely self-contained than for smaller industries, in spite of techniques such as the industrial estate.

<sup>23</sup>This argument is that associated with Hirschman; see his "Interregional and International Transmission of Economic Growth," chapter 6 of this volume. It is developed in some detail in Alonso, "Urban and Regional Imbalances in Economic Development."



may be slowed by the shift to nationally produced intermediate goods that are likely to find both their materials and their markets within the large existing concentrations, and therefore to be very much oriented toward the existing centers of production. Further, the national market for these products will probably be relatively small, so that production runs will typically be short and tailored to the particular order. As I have discussed above, such a dependence on close contacts and on externalities will lead to concentration of these industries in the existing centers. In spite of the possible effects of some sectoral shifts, however, the maturation of industries and of physical and social infrastructure argues for a natural process of relative dispersal of development to secondary centers.<sup>24</sup>

#### **Location of Industry from the Point of View of Public Policy**

Most of the considerations outlined in the preceding pages regarding private enterprise apply with equal force to projects in which the location decision is the responsibility of government. Thus it has been suggested that managers tend to overestimate the profitability of the center relative to the hinterland because of their lack of objective knowledge about conditions in the periphery, because of the relative ease of making new investments at the center, because of the strong preference for the metropolitan environment, and because of lack of interest in the periphery per se.<sup>25</sup> In this paper, taking a very similar list of considerations, it is argued that such behavior by entrepreneurs is quite realistic, using a probabilistic logic which discounts the unproven expectations of the hinterland in proportion to their uncertainty and the delays in their coming to fruition. Economic models that assume full information are unsuited to the conditions in developing countries; the lack of information should not be treated as an accidental complication. In point of fact, the lack of information is a structural aspect of underdevelopment. Data presumes records, a paper culture, a sophisticated and technically competent bureaucracy, and sufficient demand for the information from government and private decision-makers and technical staff to justify the heavy investments necessary to gather, store, and retrieve the information. These conditions are not met in developing countries.

Bryce points out that in developing countries, where the economic base is small and precarious, the appropriate strategy is one of production rather than innovation.<sup>26</sup> One of the profound disadvantages of underdevelopment is that, since resources are few, enterprises with high probability of failure must be avoided if safe although less profitable alternatives are open. One of

<sup>24</sup>A dissenting view appears in P. Hauser, ed., *Urbanization in Latin America* (Belgium: United Nations, Economic and Social Council, 1961). It is argued there that producer goods and capital goods plants tend to choose production sites close to the necessary raw materials. This view seems to slight not only the great pull of the market, but also the fact that most of the inputs into these activities are themselves manufactures rather than "raw materials."

<sup>25</sup>Hirschman, *Strategy of Economic Development*, p. 185.

<sup>26</sup>Bryce, *Industrial Development*.



the advantages of the larger size of developed countries is that it permits them to assume the risks of innovation, since enough experiments are undertaken to guarantee that the failures will be more than compensated for by the successes. In the same sense, the choice of a hinterland location in a developing country may be regarded as an innovation. Only nations whose economy has become sufficiently large and secure will be able to afford the risk of failure. Similarly, large firms consisting of several plants will be those most likely to take the chance.

The identical logic applies to the location decision by a governmental agency entrusted with one or a few projects. Rational programming at the level of the individual agency will lead to location decisions that are virtually indistinguishable from those of private enterprise. It is only when agency programming is informed by national policy considerations that the location decisions may vary. The following pages will sketch some of the problems and techniques of linking national policies and agency programs.

The national government in many countries is engaged in a great many enterprises and, by comparison to the individual agency, may have more of the advantages of size just mentioned and thus be able to afford a more adventurous policy, risking a few failures in exchange for greater overall returns.<sup>27</sup> The problem is how to transmit this overall strategy to the programming agency. One possibility is to require that agencies examine alternative locations before arriving at the decision. Such a policy is not without costs, in spite of its deceptive modesty. It may stretch very thin the technical resources of the operating agency; it may delay the initiation of projects; it may invite irrational decisions by stirring up a hornet's nest of political pressures, and it may cost a surprising amount of money. Yet this policy may be indicated if there is an expectation that enough advantages will be uncovered to compensate for these costs.

The principal reason for this possible divergence between national policy and private or agency decisions is that the national government is, or should be, concerned with national profitability rather than with project profitability. The costs of examining alternative locations compared to the probable benefits may be excessive when they are calculated for individual projects, but costs and benefits may be quite different if we consider a large number of

<sup>27</sup>Size makes it possible to take advantage of risky situations. Rich countries, large companies with several plants, and the government as the representative of the national economy are big in comparison to the single-plant enterprise and are therefore better able to chance distant locations. This has to do with the notion that, under diminishing marginal utility, a possible loss of a given size is more hurtful to a smaller actor. The first discussion of this principle is by D. Bernoulli, whose "Exposition of a New Theory of the Measurement of Risk" (1738) appears in translation in *Econometrica* 22 (1954): 23-26. He observed that "a man who risks his entire fortune acts like a simpleton, however great may be the possible gain," and that "it may be reasonable for some individuals to invest in a doubtful enterprise and yet be unreasonable for others to do so." A great deal of current interest in this line of reasoning has resulted from its presentation in J. von Neumann and O. Morgenstern, *The Theory of Games and Economic Behavior*, 2d ed. (Princeton: Princeton University Press, 1947), chap. 3 and Appendix, "The Axiomatic Treatment of Utility."

projects simultaneously. This is only one of many instances of this principle. Its basis is that the national government represents the nation and therefore should take into account the external effects of actions while the single public project typically carries its books in a manner quite similar to that of a private enterprise. In the next few pages I will mention some of the principal ways of introducing a consideration of economic costs (as distinct from money costs) into the location decision by the agency.

The most general influence upon the location decision flows from a possible difference in the interest rate applied to capital costs and the discount rate applied to future profits.

The appropriate rate upon which to calculate interest costs on public investment has been the subject of considerable recent discussion. The arguments are complex and there is no unanimity on the proper rate, but a substantial number of economists hold that a rate should be used substantially below the interest rate prevailing in the free market. Obviously, such a lower rate would favor alternatives using more capital than the alternatives chosen by free enterprise or by government using higher rates. If agencies in charge of projects were directed to base their calculations on cheap capital, it might favor distant locations when such locations required larger investment in infrastructure. Similarly, a policy that made capital relatively cheap would favor investment in plant and machinery, which, we have argued, is more typical of standard-product industries with long production runs.

It has also been argued that the rate of discount to be applied to the future stream of income should be lower for public projects than that used (often only implicitly) by private entrepreneurs. Among the arguments for this proposition is the permanence of institutions and the society they represent, compared to the briefer and more uncertain lives of individuals and private concerns. If public agencies are directed to apply a low rate of discount to future earnings, they will be less influenced by delays in the realizing of the advantages of distant locations. As has been mentioned, these delays may stem from the length of time necessary to develop the labor force, or from a lengthy period of adaptation of production processes to local conditions, or from the time it takes to establish efficient regular procedures of communication with suppliers and customers.

These three simple directives from a central planning agency to operating agencies might alter the location decisions in some projects, and it may be conjectured that they would favor to some degree hinterland locations. On the other hand, it must be recognized that in reality many developing countries are politically unstable, and that a particular government, or some officials within a government, may want to obtain even quicker results than private entrepreneurs. In such cases they will have a very high discount rate and favor projects with very quick returns. They will shun distant locations because the exploration of alternatives would delay the project and because projects built near the centers of population have greater showcase value. Whether or not such considerations are valid, they are certainly present in

many cases, and no technical means exist to evaluate the relative value of political stability.

Still, the decisions to explore alternatives and to use a low rate of interest and a low rate of discount are only a beginning toward the assessment of economic costs and benefits that reflect the national interest, as opposed to the assessment of money costs and benefits which characterizes private and most public-project accounting. Other steps can be taken, some of which are relatively simple. For instance, if one of the inputs bears a government tax, its economic cost would be obtained by deducting the tax from the purchase price;<sup>28</sup> conversely, if one of the inputs receives a subsidy, the cost of the subsidy should be added to the price. The general principle is that taxes and subsidies are transfer payments and should be eliminated from the money costs in order to arrive at the economic costs. The cost of land (exclusive of improvements upon it) presents a particularly difficult theoretical and practical problem in this respect. The purchase or leasing of land for a project is, of course, a money cost, but insofar as the transfer of money only transfers the asset (land) from one holder to another, it is not an economic cost and should not be counted among the economic costs. But it may also be argued that the price of land represents the opportunity costs that are foregone by excluding some alternative use for that land; if this view is accepted, land costs should be taken into account. Furthermore, it may be argued that the price of land, particularly if it has been unoccupied or little used, is some measure of its productivity in the intended use (say, as a location differential rent) and consequently should be counted as benefit rather than as a cost. The matter is far from clear, and I am inclined to disregard the cost of idle land either as a cost or as a benefit from the public point of view. Land that is in use should be treated as a cost to the extent of its advantage over an alternative location for the use being displaced and the costs of transferring it.

The techniques of cost-benefit calculation are only now evolving in the developed countries, and they encounter new difficulties in their adaptation to the developing ones. There are practical reasons for this, such as the scarcity of technical personnel who know how to employ these techniques, but there are also fundamental theoretical reasons. Two will be mentioned here. First, the economic cost of an input is the marginal cost of producing it and delivering it. From economic theory and ease of procedure, if not from empirical observation, it is commonly assumed in developed countries that the price of an input is the marginal cost of production. This assumption is far shakier in a developing economy, where the pricing mechanisms are more capricious, the economy less integrated, and money a less certain medium (in part because of inflation). Thus, the assumption that price equals marginal cost is less tenable. Second, because of the small size of underdeveloped economies, the calculations of cost-benefit analysis may be regarded as a

<sup>28</sup> In some cases the tax may be, in effect, a charge for services or facilities. In such cases, the tax should not be deducted.

partial differential of a general equation for national income, in which the final net benefit is the net of the effects on the various sectors of the national income of introducing a change: the project. That is, it is a sum of marginal effects. In a developed economy, with a large national income, any one project will be small in its effects and these marginal effects may normally be taken to be constant rates. But in a developing country, starting from a small economic base, a project is likely to be large and to have profound structural effects, changing, for instance, prices and rates of output. Such considerations make the applications of cost-benefit techniques in developing countries more difficult, but it may also be argued that the relative scarcity of capital makes them more necessary. In using these techniques, there is call for a greater awareness of their limitations and of the many important factors that do not enter into the calculations.

The calculation of marginal costs can be difficult, but its difficulty can be exaggerated. For instance, if the transport system is underutilized, the actual transport costs are likely to be lower than the freight charges, and an educated guess may be sufficient if an exact measurement is not possible. Or, if the transport system is overloaded, the marginal costs may be far higher. A precise evaluation may be impossible because it involves, for instance, the costs of congestion borne by others. It may be sufficient operationally, however, to use money costs inflated by some very rough estimate—or money costs may be used merely with a warning that this is an underestimate.

Far more difficult is the evaluation of external effects. For instance, the location of a particular project in some particular locale may make it possible for other enterprises, either public or private, to come into existence, and this should be counted among the benefits. Of course, the wages of labor that would otherwise remain unemployed should be counted as a benefit from the national point of view, but it may be more difficult to estimate the increase in economic activity that this increase in demand causes. The use of a multiplier is indicated, but it would be far less reliable than in a developed economy in that the response to increased demand may result in sharp price effects because of inflexibility on the supply side.

The problem of city size has been discussed by many, and it often forms a cornerstone of location policy in both developed and developing countries. In general, most nations seem to feel that their principal cities are excessively large; consequently, they follow policies of decentralization, particularly for manufacturing industry. The factual basis for this policy is lacking: no one knows when a city becomes too big. Nonetheless, this concern is so central to so many countries that some of the theoretical aspects will be discussed here. In general, it is based on a belief that some set of costs rises after cities reach a certain size. Among the costs mentioned are those of traffic congestion, water and sewage disposal, shelter, policing, and some social costs. While it must be emphasized that there is no empirical verification of the fact of rising costs, the argument, in one of its most sophisticated forms, runs thus: any particular businessman or government agency, in moving into

a congested center, will pay only the prevailing or average costs for labor, food, movement, utilities, etc. The addition to the size of the city of this new activity will change these costs only imperceptibly for any user. The marginal costs (that is to say, the addition to total costs) of the new arrival, however, will be much larger than average costs. Thus, if costs are rising with urban size, new activities will continue to be attracted to the city far beyond the point at which the increasing marginal costs begin to make this a relative or even absolute loss from the point of view of the national economy. A cost-benefit analysis would use the marginal costs produced by the external diseconomies of congestion, etc., in evaluating that location. If we knew what these costs were, the central planning agency might require operating agencies to use them in their calculation and the government might charge private businesses the full costs by some form of taxation.

It was pointed out in an earlier section that there are reasons to believe, however, that productivity may also increase with city size. Among the factors mentioned was the facility of communications and the availability of information, the availability of specialized services, the richness of inter-industry linkages, and the principle of massed reserves. If productivity increases with urban size, then the project or the businessman will take into account only the effects of these externalities upon his firm, which in a sense may be termed the average productivity. But the project will not take into account, under normal procedures, the external economies that it produces for other activities. In other words, the symmetrical analysis should be performed on the benefit side as on the cost side, and the public agency should take into account the marginal effect on total production in the city. If productivity is rising with size, marginal product will be considerably larger than average product. No more do we have reliable information as to whether productivity rises with urban size, but a great deal of corroborating evidence suggests that it does. The magnitudes involved in terms of per capita income, net regional income, value added per industrial worker, local government expenditures, and cost of living indexes for those few countries in which they are available, suggest to this writer that marginal product is far greater than marginal cost for large cities, and consequently that antiurbanization policies are in error with respect to the goal of national production, for they compare marginal cost to average product.

The external benefits of a hinterland location are even more elusive. The contribution of the project to the external economies of the location must usually be estimated not for an existing center of economic activity, but for a future one to which the particular project will contribute. Clearly, if we find it difficult to evaluate the external effects in cities that exist, it will be even more difficult to do so where the development is yet to come. This is particularly true if we believe that there is a minimum critical threshold or takeoff size for a growth pole. Current models used to calculate growth in a local economy as a result of new industry, whether simple multipliers or more complex input-output or money-flow models, might be called incremental



growth models and do not lend themselves to the study or prediction of the quantum jump implied in the concept of critical size.

Other important considerations do not lend themselves easily to measurement. Three will be mentioned here. First, the project may contribute to the transformation of society in the less advanced parts of the country to new attitudes, new awareness, and new patterns of behavior better suited to economic advancement. In other words, a hinterland location may be viewed as an investment in human resources which may be of greater significance than the located project itself. Second, the location of a project in the hinterland will usually promote knowledge about that locality and help integrate it into the information web of the more advanced sectors of economic activity. The increase of information will reduce the discount arising from uncertainty which is applied to other possible projects at that location. In so doing, it lowers the threshold at which local opportunities become attractive, and several new enterprises may become feasible. Third, it was mentioned earlier that most developing countries, especially those with a colonial background, feature their most important development in coastal cities. In some of these countries a turning inward, toward the "empty" hinterland, may have significance as a symbolic act of national identification, creating a new frontier to call forth the enthusiasm and energy of the people.

In the preceding discussion we have focused on the goal of national economic growth, but many nations also want to move toward interregional equality of incomes and are willing to sacrifice some absolute growth to this purpose. A project may be located in a backward region for this objective even if project profitability (in terms of the internal accounts of the project) and national profitability (in terms of contribution to national economic growth) are greater in the city. There is, of course, no formula to study the trade-off between national growth and national equality, and the decision is ultimately a political one. It must be observed that, because the choice is a hard one, too often facts are twisted to make it appear that in all cases both objectives are served simultaneously, whereas in all likelihood they are often at odds.

The objective of interregional equality has received relatively slight systematic consideration in the literature.<sup>29</sup> It is a particularly difficult subject to discuss in nontechnical terms because the measures of equity tend to be more technical and far more dependent on the geographic definition of regions than are those of national growth. Further, since population is commonly free to move from one region to another, a clear distinction must be maintained between the geographic unit and the people involved. Thus, as has been mentioned earlier, a project in a backward region that depends for its labor on people brought in from outside the region may raise the per

<sup>29</sup>A good review of some of the issues and much of the literature is available in T. Reiner, "Sub-National and National Planning: Decision Criteria," *Papers of the Regional Science Association* 14 (1965).

capita income of the region but lower the welfare of the original inhabitants if it introduces shortages or higher prices. Similarly, a project located outside a region, by draining off some of its surplus labor, may in fact raise per capita incomes both for those who leave the region and those who remain behind. Further, a policy of developing growth centers in backward regions may (or may not) decrease the inequality of incomes between regions but increase them within regions. It is unclear in many cases whether the objective is to decrease inequality of income distribution for the national population as a whole, with the region serving as an instrumental concept for the guiding of policy, or whether the regions are viewed as organic entities so that the equalizing of incomes among regions is itself the objective. A full discussion of these issues is clearly beyond this paper and is unavailable elsewhere. Only a few observations will be made.

In mixed economies, the use of lower rates of interest, government loans or insurance of capital investment, or direct investment by the government are among the common devices to induce industry to locate in the hinterland. It should be noted that there is a certain irony in the use of these inducements, for they will operate most strongly for industries that are capital-intensive, and, if technical substitution is possible, it will encourage them to substitute capital for labor since they make capital cheaper. Such industries will tend to rely on skilled labor, which will often be brought in from outside the region, and will usually be large-scale industry, which, as has been discussed, tends to be more self-contained and therefore less likely to induce local growth because of the sparseness of local linkages. Only recently have a few countries, such as Great Britain and the German Federal Republic, begun to experiment with labor subsidies as inducements to attract labor-intensive industries. Insofar as the reasons for encouraging such locations are to put local surplus labor to work and to discourage migration to the cities, it would seem reasonable to try to attract industries for which the labor factor is preponderant.

The costs of such subsidies, or of direct investment in infrastructure, or of government services to attract industry to poor regions, should be attributed to the equalization goal if the industry would have come into existence elsewhere within the country in the absence of the inducements. They would be the costs of making what may be termed a geographic transfer payment, and they would pertain to redistribution rather than to national production. The benefits that derive from the industry should not be counted in the national profitability calculus of cost-benefit. They pertain to regional profitability, which is as distinct from national profitability as project profitability is. Only if the locational inducements cause industry to be created within the national territory which would not otherwise have come into existence do regional and national benefit calculations coincide, and even then secondary effects may differ at these levels if the location of industry shifts secondary activity from one region to another. It may be noted that the common confusion between equalizing and national product objectives is based on proceeding as if re-

gional benefits were identical with national benefits, and ignoring the possibility that we are taking from one region to give to another.

None of this must be interpreted as a disparagement of the equity goal as a valid criterion for the location of industry from the point of view of public policy. The point being made is that the analysis with respect to this goal is too often poorly handled.

### **Regional Policy as a Location Factor**

Regional policy is not commonly regarded as a location factor in its own right. Rather, regional policy is viewed as the basis of particular programs and actions that modify other factors. Thus, capital may be made cheaper in certain locations by preferential rates, or the disadvantages of a distant location may be mitigated by a subsidized transport rate, or, in more extreme cases, certain locations will be forbidden and others may be made mandatory by government decree. But regional policy in itself may operate as a location factor in developing countries, over and above particular government actions to modify the other location factors.

Much of the preceding discussion was based on the failure of the conditions assumed by classical location theory to match the conditions in developing countries. Among the divergences are the following: (1) in developing countries there is a general scarcity of information and a greater scarcity in the less developed areas of those countries; (2) in a country undergoing development, the modern economic base will be relatively small; therefore, changes are likely to be relatively swift and radical, so that there is more structural fluidity and less validity to the traditional assumption of "all other things being equal" of the traditional theory; (3) the shortage of entrepreneurial, managerial, and technical personnel in and out of government is one of the principal constraints upon development.

Given these departures from the world of classical economics, the very existence of a regional policy may affect the location of some projects. For instance, an announced and publicized decision to develop some provincial center may call the entrepreneur's attention to that center as a possible location independently of the concrete steps taken to develop it, whereas if no center were designated by the government the entrepreneur would not know where to look in the provinces and he would be more likely to settle in the principal cities. If there is some degree of confidence in the constancy and efficacy of government action, the general statement of regional policy serves for the analysis of projects to reduce uncertainty about the structural changes that may be expected in the nation's economic landscape, thus lowering the discount applied to expected returns to allow for uncertainty. Of course, the simple fact that regional plans normally contain a collection of information about the region will serve to increase available knowledge about it. Thus, the regional plan will play a role in reducing uncertainty about present circumstances and future structural changes.

This suggested effect of a regional plan is in some ways similar to the



French concept of indicative planning: when diverse activities have common interests, it is sometimes sufficient to point to a common goal, without need for either commands or special inducements. Imagine that several projects are about to be undertaken, some by the government, some by private concerns. In the absence of a plan, some will locate in the principal city and some will locate here and there in the interior. Designation of a locality as a growth locality may inform several of these activities of the opportunity of coming together and benefiting from mutually induced externalities.

It must be understood, however, that such a function of the plan does not absolve the government from a responsibility for concrete action or from the need to study and prepare for the complementarity of needs of the participating activities, including governmental investment in infrastructure.<sup>30</sup>

The indicative function of the regional plan is not without problems. The plan itself will be drawn in the midst of uncertainty as to what is possible or even desirable, and will probably need frequent revision as more is learned about actual circumstances and opportunities. The success of the indicative function, however, depends in large measure on the confidence of the participants in the stability of the plan and the commitment to specified objectives. If these change frequently, the utility of the plan as an instrument to reduce uncertainty will be negated.

The shortage of entrepreneurs, managers, and technicians represents a narrow national capacity to collect and evaluate information. The common injunction that location decisions in each case be made after a careful examination of a multitude of variables is not realistic. Alfred North Whitehead has observed:

It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.

Thus, one of the principal functions of regional planning may be the economizing of scarce project-planning energy. It achieves this by indicating prob-

<sup>30</sup>Among regional planners, balanced regional growth usually means either that all regions grow at the same rate or that the slower regions grow faster to catch up with the more developed ones. It has recently been pointed out that the "big-push" sectorially balanced growth strategies of Rosenstein-Rodan and others, which stress the complementarity of certain sectors, is likely to result in sharp differences in regional rates of growth, with a bias toward the more developed centers, thus leading toward greater regional inequality and "unbalanced" regional growth. See W. F. Ilchman and R. C. Bhargava, "Balanced Thought and Economic Growth," *Economic Development and Cultural Change*, July 1966. In the terms developed above, balanced sectoral growth is oriented to the goal of national growth, and balanced regional growth to the goal of equalization.

able decisions for classes of projects and by foreclosing other decisions. A plan may indicate what types of industry may decentralize and what types should continue to concentrate; it may indicate the likeliest locations for industries with certain characteristics; it may insure rates of return on locations that suit certain general objectives, etc. Mention has already been made of indicated rates of interest and discount which might be made to apply for public projects. Similarly, shadow prices<sup>31</sup> or techniques for computing them may be indicated by a central agency for some commonly used inputs so that these do not have to be recalculated by each project that undertakes a cost-benefit analysis. And, of course, certain information can be provided and distributed, including projections of basic variables for general use.

Two aspects may be distinguished in the use of regional planning to economize on the demands placed upon the managerial and technical resources. The first is the production and distribution of information which can be used by more than one project. The second is the ability to take an overall view of the evolving spatial structure of the national economic landscape in the process of development. Overall patterns may be discernible in the whole which are not apparent in the individual examination of projects. Examples of this are information on the success or difficulties of certain types of locations for certain types of industries, or a long-term perspective for the growth of particular localities such as secondary centers in the interior, or aggregative projections of traffic volume on certain arteries which can indicate when saturation may be expected. If a long-range view is taken of the economic macrogeography of a developing country, it is certain that there will be fundamental structural changes, although there normally will be considerable uncertainty about the precise form of these changes. Project analysis, on the other hand, tends to work within a constant structural matrix while focusing in greater detail on the specifics of the action being considered and to use short-run curves. The rapid change of the economic matrix when development occurs makes the use of short-run curves and partial analysis less valuable for developing countries than for developed ones.

In the absence of limitations on technical resources, both the aggregative macrogeographic and the detailed project studies should be carried out in full. But since sharp limitations exist, the aggregative analysis can usefully substitute up to a point for the detailed analysis, and a mixture of these approaches will be more fruitful than exclusive concentration in either one. If only detailed project analyses are carried out, their joint effects and their longer-run consequences will not be anticipated, and important opportunities will be missed from lack of awareness of broader trends. If only the generalized analysis is pursued, many projects may have disastrous results from having overlooked important details, or equally important, the nation may

<sup>31</sup>Economic costs when they differ from money costs.

find itself with excellent general policies, but without concrete programs or specific projects to give substance to the policies.<sup>32</sup> In short, the lack of information in specific cases, the lack of technical personnel, and the rapidity of change in developing countries place a special premium on rules of thumb, general guidelines, and policy.

Although profound changes are certain in the structure of the economic landscape of the nation, our present knowledge is too tenuous to permit anything but guessing as to the nature of these changes. It is necessary to make these guesses, but new information and the fact that any vision of the future is a compound of judgment and opinion will result in a tendency to instability of regional plans. It must be reemphasized that frequent changes will reduce the importance of regional policy as a surrogate for information by undermining the confidence of project planners in the validity of that vision of the future. At the same time, when new and better information and fresh insights become available, new visions of the future will arise. Thus there is a tension between the function of the plan as a stabilizer in an excessively fluid situation, and the logic of continuing planning, feedback, and revision in the light of increasing understanding. No easy solution presents itself to this dilemma.<sup>33</sup>

It may also be noted that, insofar as regional plans both produce and substitute for information about the national economic landscape, they will be of greater influence as a factor of location for mature industries than for industries that are getting under way. The latter still face uncertainty as to processes, relations to complementary activities, and other factors that call for adaptability rather than predictability, and therefore they will continue to favor the principal cities. On this basis, it may be hazarded that the influence of regional policy per se as a factor of location will be strongest in the middle stages of economic takeoff, after the beginnings but before maturity, when the patterns have become clearer and information a matter of course.

In conclusion, two general points must be made in regard to national regional policy. The first is that in geographic terms, economic development may be viewed as the integration of the economy over the national territory. By this is meant the elimination of irrationalities in the production and exchange of goods and services throughout the country, overcoming ignorance, prejudice, and inertia, and facilitating the means of exchange among diverse parts of the nation.<sup>34</sup> In this respect, the nation must look not only to the production and even the immediate external effects of any project, but also to

<sup>32</sup>This vital point is documented in A. Waterston, *Development Planning: Lessons of Experience* (Baltimore: Johns Hopkins Press, 1965).

<sup>33</sup>A particularly thoughtful discussion of the relation of scientific knowledge and the exigencies of the actual situation in regional planning is available in J. R. P. Friedmann and W. Stohr, "The Uses of Regional Science: Policy Planning in Chile," *Papers of the Regional Science Association* 20 (1967).

<sup>34</sup>See Balassa, *Theory of Economic Integration*, and Friedmann, *Regional Development Policy*.

its modernizing influence upon the local population and its force as an integrator of that part of the nation and its human and natural resources into the fabric of the modern nation that is trying to come into existence. Needless to say, there are no quantitative or other scientific methods to evaluate the trade-off between economic efficiency of a project and its long-range effect in the making of a modern nation.

The second has to do with the fact that in most countries, regional planning is done by the executive branch, and budgetary and perhaps substantive approval is reserved to the parliamentary branch of government. The executive branch is divided principally into sectoral or functional divisions such as ministries. The legislative branch, almost without exception, represents territorial divisions. The technical planning function is usually housed within the executive branch, and most national plans thus far have been sectoral plans, formally or less formally representable in an input-output or some other social accounts table. The emergence of national regional planning in the past few years presents a formidable problem: as a technical matter, it belongs to the executive branch; as a territorial decision favoring or denying growth, it becomes of immediate and concrete importance to individual legislators. It is therefore clear that the function of regional planning, in contrast to that of sectoral planning, will be continually disrupted and enriched by the introduction of local voices into the rumblings of professional technical discourse.

### Summary

The location of industry has been discussed principally in terms of a choice between a principal city and hinterland locations. Classical location theory was reviewed briefly but it was stressed that, even within the terms of this theory, the objective of minimizing transport costs is a poor substitute for some real objective, such as maximizing of profits.

However, the assumptions of classical location theory relating to unlimited managerial resources, full information, predictability of the future, and "all other things being equal" are unsuited to developing countries. After consideration of the form of the transport network, we discussed the consequences of distance as time, of the personal space preferences of management, of lack of information about hinterland locations and the pressures for rapid development, and of some aspects of the labor force at hinterland locations. A broad definition of external economies included consideration of the principles of multiples, massing of reserves, and bulk transactions, relating them to city size, as well as the effects of the ease of communication, of national customs, and of the lack of standardization of procedures. These common considerations for the location of industry in developing countries were compared to those for infant industries in developed countries. It was suggested that much of the apparent overconcentration of development in new nations is a functional aspect of the early stages of development.

Location of governmental industry will not differ significantly from that of

private industry when only one project at a time is considered and calculations are based on money costs. It was suggested that directives to project managers to examine alternative locations and to use lower rates of interest for capital and lower rates of discount for returns might result in more hinterland locations. Some further considerations were outlined for evaluating location in terms of economic costs in the framework of cost-benefit analysis. It is unclear whether these considerations will alter the net balance of locations away from the principal centers, although it may change the locations of some industries. Reference was made to the public objective of interregional equalization as distinct from that of aggregate national growth.

Finally, regional policy per se as a location factor was considered as a substitute for and generator of information and as a useful device to conserve and stretch scarce technical manpower. Regional policy may also affect the logic of location decisions when it gives consideration to longer-run effects and the economic macrogeography of development.

---

## 4 The Nature of Economic Regions

August Lösch

Impressed by the accidental way in which states are created and smashed, we are looking out for a more natural and lasting spatial order of things. Geographical and cultural regions, however, are from an economic point of view just as artificial units of reference as states are. True enough, they all are of some economic relevance, but this does not alter their essentially non-economic nature. Important as their balance of payments, their price levels, their barter terms of trade may be for them, to *us* these averages and aggregates are entirely arbitrary and accidental. It is independent economic regions that we here discuss, regions not derived from but equivalent to those political, cultural, geographical units.

Even if we already knew the characteristics of economic regions—which we do not—their counterparts in the world of reality would be likely to differ more from each other than from an ideal picture. Hence studying the ideal region is both the only way to learn about the *essential*, and the first step towards investigating the *actual* structure of any real economic region. So we shall deal first with the theoretical nature of such regions, and second with their actual existence.

### I

Let us start from very radical assumptions in order to prevent any spatial differences of an uneconomic origin from hiding in our starting points. We assume a vast plain with an equal distribution of raw materials, and a complete absence of any other inequalities, either political or geographical. We further assume that nothing but self-sufficient farmyards are regularly dispersed over that plain. How can any spatial differences possibly result from this initial situation?

Supposing one of those farmers tries to produce a certain commodity beyond his needs, will he be able to sell the surplus? He will be helped by the economies of large scale production, and handicapped by costs of transportation. Will the balance be in his favor? If his neighbors all have a similar way of living, the demand curve of one of them will be typical for the others as well. Let us assume  $d$  in Figure 1 to be such an individual demand curve for beer.  $OP$  being the price at the center of production  $P$ , the demand of the people living there will be  $PQ$ .  $PR$  being the freight from  $P$  to  $R$ , the demand of each of the people living in  $R$  is  $RS$ . Farther out, at  $F$ , where the freight is  $PF$ , no more beer will be sold. Hence  $PF$  is the maximum shipping radius for beer, and the total demand within that radius is equal to the volume of the cone which we get by rotating the triangle  $PQF$  around  $PQ$  as axis. Figure 2 shows that cone. To repeat: its volume, corrected for the density of population, is equal to the total possible demand if the price at the factory is  $OP$ . For other prices at the mill we get other cones of demand, and as a final result the curve  $\Delta$  of Figure 3, that represents the total demand as a function of the price at the mill.  $\pi$  of Figure 3 is a so-called "planning curve," showing the minimum costs at which a given output could be produced if a new factory had to be built for that purpose. Only if the planning curve  $\pi$  intersects or is to the left of the total demand curve  $\Delta$ , is it possible for our farmer to run a brewery. Otherwise he would produce at a loss.

The shape of a trading area, however, is not a circle, as we have so far assumed. For even if the whole country were filled up with such circular areas that are close enough to just touch each other, a number of people could still successfully try to enter the brewing business. For all the black corners in Figure 4 are left unused, and moreover, as has been shown by Chamberlin,<sup>1</sup> the size of the

<sup>1</sup> For those not acquainted with Chamberlin's theory it may be worth while to point out that his argument is based mainly on two facts: (1) Due to product differentiation, of which differentiation of the seller's location is just a special case, the demand curve facing the individual seller is not horizontal (as in pure competition where the product is perfectly uniform) but has a negative slope. If e.g. the seller raises his price, not all his customers will buy from his competitors as in a perfect market. To a number of them the special advantages (e.g. of convenient location) offered by him will be worth the higher price. (2) As long as the demand curve is to the right of the cost curve the extra profits thus possible



individual firm will be reduced from  $MN$  to  $M'N'$  (in Figure 3) without rendering it unprofitable. The way to make use of the corners is to change the shape of the area into a regular hexagon. This will shift the curve  $\Delta$  slightly to the left, as the hexagon is somewhat smaller than the circle that circumscribes it. Moreover, by Chamberlin's operation the size of the hexagon will be reduced until it is so small that the corresponding demand curve  $\Delta'$  just touches the offer curve in  $N'$ . Now apparently no more people can enter the brewing business.<sup>2</sup> As the largest possible shipping radius results in a total demand  $MN$ , so the necessary minimum radius must yield the demand  $M'N'$ . Figure 4 shows the development from the largest to the smallest possible shipping range.

Two other possibilities of avoiding black corners are conceivable, namely the square and the triangle. But it can be shown<sup>3</sup> that the hexagon has an economic advantage over both: it affords the larger demand per square mile, provided the total area is the same in all cases. *The hexagon is, therefore, the most economical shape for trading areas.* For every commodity, a trading area in the form of a hexagon with a characteristic inner radius  $\rho$  is necessary and sufficient to render the production of this commodity profitable.

The trading areas of the various products look like nets of such hexagons, from very small ones to very large ones, depending upon the product. We can throw these nets over our plain at random. In spite of the resulting disorder, every place on the plain would have access to every product. Several considerations, however, which can only be mentioned here, suggest a more orderly and at the same time more economical arrangement. In the first place, we lay our nets in such a way that all of them have one center of

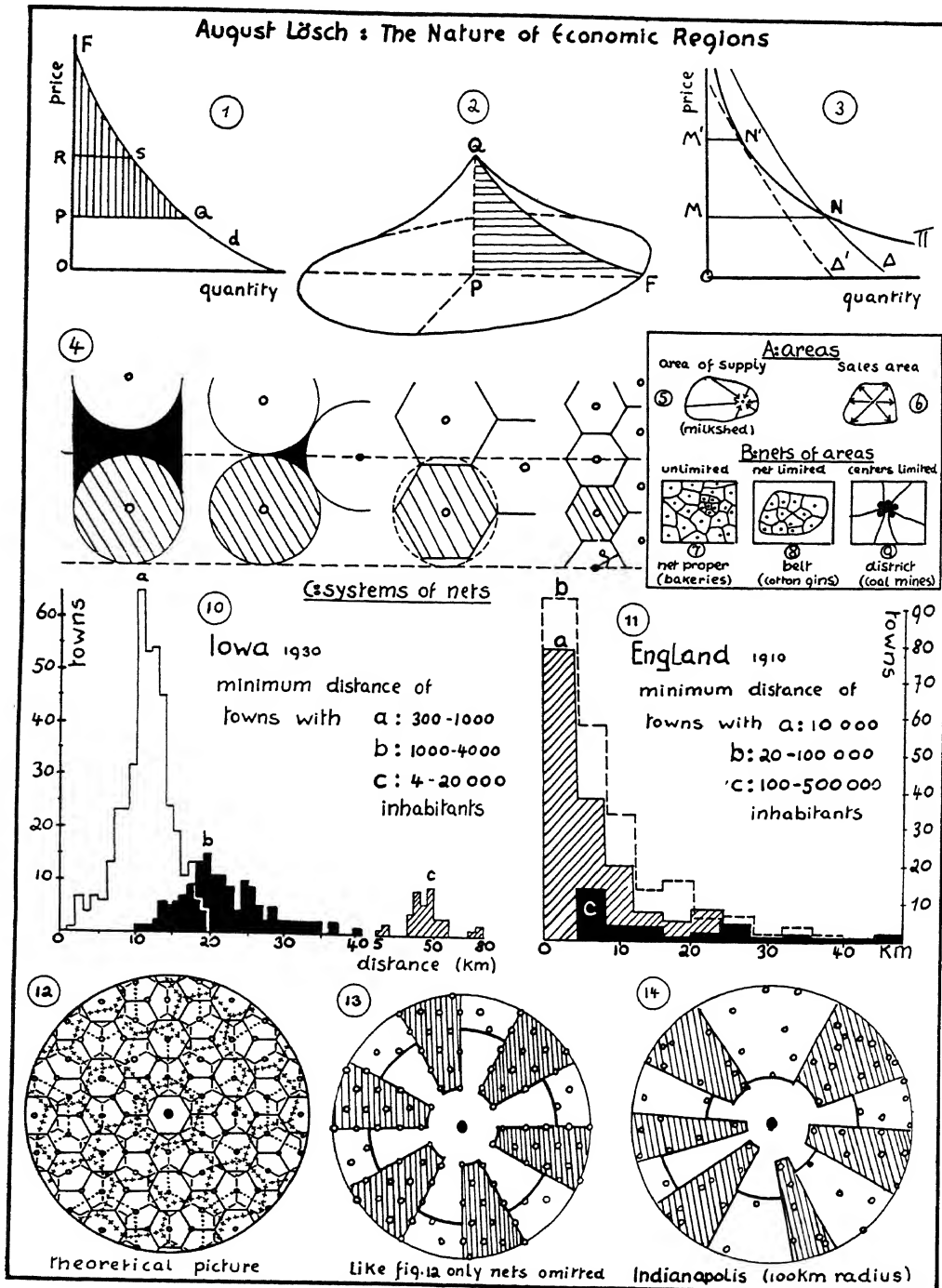
---

will attract new competitors. They will sell products slightly different from those already in the market, or, as in our case, locate their businesses at places more convenient for part of the buyers. This will shift the demand curves of the old establishments to the left until they just touch the cost curves and all extra profits are wiped out. (See E. Chamberlin, *The Theory of Monopolistic Competition*.)

<sup>2</sup> We disregard here the possibility of reducing the area even more through spatial price discrimination.

<sup>3</sup> Whilst a more accurate and detailed proof is too lengthy for this short paper, the plausibility of our assertion can readily be seen from the fact that the regular hexagon has the advantage over the circle of using up all the territory, without departing as far from the ideal circular shape as either square or triangle.





production in common. This point will enjoy all the advantages of a large local demand. Secondly, we turn the nets around this center so that we get six sectors where centers of production are frequent, and six others where they are scarce, as is shown on Figures 12 and 13. This arrangement does not deprive any place of its access to every product, and at the same time provides for the best lines of transportation. It can be shown that the aggregate of freights is a minimum,<sup>4</sup> and the final result is a complicated but orderly system of market areas. How many of these self-sufficient systems will come into existence on our plain depends merely upon the commodity which has the largest necessary shipping radius, as long as there are no economic limits to the size of the central city.

More striking about our result than any particulars is the fact that we suddenly have crowds of economic areas on a plain which we deprived of all spatial inequalities at the outset. We first have the hexagonal market area surrounding every center of production or consumption. Second, we have a net of such areas for every commodity. And third, we have a systematic arrangement of the nets of market areas of the various commodities. It is the latter, the self-sufficient system of market areas as shown in Figure 12, that I should like to call the ideal economic region. How much of it we find in reality will be discussed in the second part of the paper.

## II

As soon as we drop the assumption of a uniform plain, the size and shape of our market areas evidently become irregular. Moreover, if we no longer stick to the supposition of a uniform product, the individual areas for the same line of production overlap, and may consequently be full of holes particularly near the periphery. Yet there are numerous instances left where our assumptions are roughly fulfilled and where our results, therefore, must hold true without much modification, as factual investigations indeed seem to indicate.

<sup>4</sup> As more centers of production coincide more consumers are able to buy from local mills than under any other arrangement of the nets. Not only the mileage of transports but the mileage of lines of transportation as well is reduced.

Actually it is not quite accurate to compare the numerous market areas of a commodity to a net. Due to the overlapping just mentioned they often rather resemble fish scales or an irregular layer of slabs of slate. In spite of this modification the essential characteristics of a net are mostly retained, and as a matter of fact most of the maps showing trading areas that were prepared either by scholars or by business men do not give any consideration to the overlapping at all. Far more important than this modification of the structure of our nets are the changes in their extension. In some instances, for which bakeries may serve as an example (Figure 7), the nets still cover the whole territory under consideration. In fact, a survey made by the author of about half the American industry<sup>5</sup> would seem to indicate that the importance of this type of production is rather underrated. Nevertheless, the very nets or at least their centers are often compressed on a relatively small space, and we may speak then of belts and districts respectively. The former case may be exemplified by the net of the areas of supply of the cotton gins that is naturally limited by the cotton belt. And an illustration of the concentration of the centers of production only are the mines in a coal district (see Figures 8 and 9 respectively). Instead of tracing out the areas, which is a very difficult task, we can show their character just as clearly by measuring the minimum distance of their centers from each other. This is done in Figures 10 and 11, not for centers of a single production but for towns of a supposedly similar economic function. In Iowa, with its rather equal distribution of production, the distances between towns increase with their size, just as in our theoretical picture based on assumptions approximately fulfilled in Iowa. In England, on the other hand, the cities cluster in the coal districts and show the same distance from each other irrespective of size. Such concentrations of the nets or their centers may have purely economic reasons such as the advantages due to the proximity of many establishments of the same branch. But it may also be a reflection of the limited geographical extension of factors of economic consequence although not of economic nature. It is worth noting, however, that these non-economic

<sup>5</sup> It will be included in a forthcoming book on the laws of location.

factors and their economic reflections are not co-extensive. For instance, the area where cotton *could* be grown is larger than the actual cotton belt.

In addition to the limited size of the nets, and the overlapping of the individual market areas, a third deviation from the ideal pattern is worth mentioning. In our theoretical deduction we had to cope with the problem of how the various nets should be located, while the distribution of the centers of production within a net was conspicuous for its regularity. Actually this too is a problem, and a very difficult one at that. Neither of the two traditional instruments of determining the geographical distribution of production can solve it: the theory of location proper cannot because it is applicable only to a single establishment, not to a whole industry; and the theory of comparative costs fails because it is applicable only to trade between men, not between countries. The only adequate solution of the location of all the interdependent centers of production is a system of locational equations which the author hopes to present later.

The systems of nets come off worse in the real world than either the nets or the individual market areas. It is simply impossible to arrange all the irregular nets in such a way that they have at least one point in common. There exists nowhere either a city with a complete set of industries or a self-sufficient region. But this is not the worst. We could at least imagine and probably find a few actual cases where regions trade their specialties with each other through their central cities, and through them alone. In such an instance a systematic arrangement of towns as in our ideal region would still be conceivable. Actually, however, small places which in every other respect entirely depend upon neighboring cities are the centers of large market areas. As far as their particular products are concerned, even metropolitan cities or the whole nation may be tributary to those little places, the industries of which neither need nor attract a large local market. Furthermore, while the regional system of nets of market areas centers in a large city, not every big city dominates such a system. Many mining towns, for instance, have not much of an economic function towards their hinterland. In contrast with such special-

ized cities, a regional center is characterized by a variety of production and trade that links it to the surrounding country. If, now, we disregard all the market areas of the type just described, a substructure of economic regions is left. They *differ* from the ideal pattern in the important respect that there are not self-sufficient; they *correspond* to the ideal inasmuch as they too are based (1) on the advantages of a large local concentration of production, consumption or trade; (2) on the most economical layout of lines of communication.

This regional substructure can be discovered almost everywhere but it is not everywhere of equal importance. Its importance can be measured by comparison with those market areas that have to be eliminated from a regional analysis as was just pointed out. To give some examples: regionalism prevails in southern Germany.<sup>6</sup> The distribution of the undisputed regional centers: Frankfurt, Nürnberg, München, Zürich, Strassburg, with Stuttgart in the middle is very regular. There should be one more center to the south of München but the Alps make this obviously impossible. The rise of München over Augsburg that had the advantage of an earlier start is worth noting. München has the better location from the point of view of our theory. It is right in the middle of the region, and at the proper distance from the neighboring centers. The German Ruhr district, on the other hand, hardly displays any regional pattern whatever. According as the systematic or the chaotic distribution of the nets of market areas prevails in a given case, we may stress or disregard the regional substructure. From this it follows that while the regional concept will be most realistic with respect to some parts of a country, it would be difficult and not very useful to divide a state up into its regions.

Finally, as to the relation between economic and other regions, it is essential for the regional system of market areas to have a center. In rare and particularly fortunate cases these economic centers are the same time cultural and political ones, thus becoming the true heart of their region, as Paris is for France.

<sup>6</sup> This has very ably been shown by Walter Christaller, *Die zentralen Orte in Süddeutschland*.

## III

To summarize, we found three main types of economic areas: simple market areas, nets of such areas, and systems of nets. Or, if we want to give a popular name to each, we may speak of markets, belts, and regions. In this sequence they become more complex, more self-sufficient, and unfortunately less real. On the one end there are the individual market areas, most simple, most real, and most dependent upon trade. The systems of market areas, or regions, on the other hand, are very complex; in an ideal case quite self-sufficient, but harder to find in reality. Many commodities are produced and traded outside of any system. And whatever systems we do find, overlap even more than the market areas of a single commodity. A clear economic region is a fortunate accident rather than a natural subdivision of states. Still, beneath a sphere of irregular market areas, we find a regional substructure of varying importance almost everywhere. Between the simple area of sale or supply and the full regional system is the net. The geographical extension of these nets or of their centers is often small. In this case these belts or districts of production or consumption are very conspicuous, but should still be distinguished from regions. A region is a system of *various* areas, an organism rather than just an organ.

---

## 5 The Geography of the United States in the Year 2000

Brian J. L. Berry

---

**ABSTRACT.** There are increasing demands that research in geography should respond to the needs of public policy. It is therefore essential to attempt to monitor geographical change, to identify its essential properties, and to understand the geographies that are most likely to emerge in the future with and without public intervention. This paper analyses the increasing contemporary polarization of the United States into a limited number of growing *daily urban systems*, expanding through continued innovation and diversification, and the *inter-urban peripheries* whose economies and populations are declining, except where minority groups such as American Indians have exceptionally high fertility rates.

In determining what is critical to further transformation of the geography of the United States, however, it is concluded that: (a) migration of the minority-group poor from the peripheries to the cores of the central cities, and (b) a resulting acceleration of the outward movement of upper-income white population from central city to the expanding outer edges of the daily urban systems, now 80–160 km away from the city centres, will invert the geography of the country by the year 2000. This tendency to inversion, supported by rising real incomes, improved highways, and the search for superior low-density residential amenities, will be further advanced by new electronic technologies that replace movement of persons by movement of messages, thus reducing and eventually eliminating the traditional role of the CBD in permitting face-to-face contacts. The coming era of *telemobility*, in which mechanical environments will be replaced by electronic environments, will push the emerging inversion of American geography into its ultimate dispersed forms.

---

It is a sign of the times that increasing numbers of geographers are to be found advising governments. The country, the agency and the particular problem may differ, but the questions asked are always, fundamentally, the same: What will be the consequences of doing A as opposed to B? What is the best way of achieving X or Y? To what extent are C and D achieving P and Q, and are unexpected or undesirable consequences emerging? To answer presupposes skills to which few among us can justifiably or should willingly admit, the ability to outline alternative future geographies, to monitor geographical change and to measure the degree to which geographical change achieves its goal. What, for example, is the most probable future geography of the United States, or some part of it, at some specific future time, in the absence of public intervention? What are the likely geographies resulting from alternative public programmes? And what is the most desirable future geography, that which most nearly achieves socially-accepted goals?

Our science may not yet be equal to the task, but I submit that, unless we develop the capacity to respond to such questions, we will find our personal alternatives vanishing as governmental support is shifted to those who can demonstrate the skill. For I accept that we are now part of an increasingly universal and post-industrial society characterized by a

trend towards increasingly sensate cultures and meritocratic élites, a society in which change, based upon scientific and technical knowledge, is guided increasingly by the public sector through its financing of research and development. I conclude that if we, as geographers, fail to perform in policy-relevant terms, we will cease to be called on to perform at all.

One of the most pressing public debates in the United States today concerns the development of a national urban growth policy. Therefore, in this paper I propose to describe what I consider to be the more salient features of the geography of the United States in the 1960s, to examine the processes that have given rise to this spatial organization, to explore the innovations producing changes in the processes and transformations in the structure, and to speculate about the probable geography of the nation in the census year 2000.

#### PAST FORECASTS OF FUTURE GEOGRAPHIES

The first step is to determine what is salient in the geography and in the processes of geographical change in the United States today. I first looked back, albeit unsystematically, to see if there were lessons to be learned from those who, in the past, have been perceptive of the future. I found a variety of examples, such as Robert Vaughan, who in 1843 wrote (pp. 90–91, 100) that ‘cities . . . must soon become . . . the acknowledged wealth and power of nations . . . the new and speedy communications which will soon be completed between all great cities in every nation . . . will necessarily tend to swell the larger towns into still greater magnitude and to diminish the weight of many smaller places as well as of the rural population generally . . . everywhere we trace this disposition to converge upon great points . . . the great tendency of modern society is toward the formation of great cities’.

Almost as an echo comes H. G. Wells’s reflection 60 years later (1902, p. 39): ‘the broad features of the redistribution of the population that . . . characterized the nineteenth century . . . [were] an unusual growth of great cities and a slight tendency to depopulation in the country. The growth of the great cities is the essential phenomenon.’ Clearly, Vaughan had identified the tendencies latent in his day.

Wells’s own predictions are still my favourites, however, for I suppose that the tasks I have identified call for fictional insight and analytical skill. We might dwell upon his *Anticipations* concerning the probable diffusion of great cities in some detail, because Wells was right in the changes he forecast, but he saw them as coming by the end of the twentieth century, failing to estimate the accelerating pace of change. What he described was not the geography of the United States in 2000, as he thought, but that of 1960. On the other hand, like Vaughan, he attached the greatest significance in his forecasts to those factors compressing time and space, and I suggest that it is for this reason that both he and Vaughan were correct.

In 1902 Wells advanced the proposition ‘that many of [the] railway-begotten “giant cities” [of 1900 would] reach their maximum in the coming century [and] in all probability they . . . are destined to such a process of dissection and diffusion as to amount almost to obliteration, so far, at least, as the blot on the map goes, within a measurable further space of years’ (pp. 45–6). ‘These coming cities will not be, in the old sense, cities at all; they will present a new and entirely different phase of human distribution’ (p. 47). ‘The social history of the middle and later thirds of the nineteenth century . . . all over the civilized world is the history of a gigantic rush of population into the magic radius of—for most



people—four miles, to suffer there physical and moral disaster less acute, but, finally, far more appalling than any famine or pestilence that ever swept the world . . . But . . . these great cities are no permanent maelstroms. New forces, at present so potently centripetal in their influence, bring with them, nevertheless, the distinct promise of a centrifugal application that may finally be equal to the complete reduction of all our present congestions. The limit of the pre-railway city was the limit of the man and horse. But already that limit has been exceeded, and each day brings us nearer to the time when it will be thrust outward in every direction with an effect of enormous relief.

'So far the only additions to the foot and horse . . . are the suburban railways. . . . The star-shaped contour of the modern great city, thrusting out . . . knotted arms of which every knot marks a station, testify . . . to the relief of pressure thus afforded. Great towns before this century presented rounded contours and grew as puff-ball swells; the modern Great City looks like something that has burst an intolerable envelope and splashed . . . the mere first rough expedient of far more convenient and rapid developments' (pp. 50–51).

'We are . . . in the early phase of a great development of centrifugal possibilities. . . . A city of pedestrians is inexorably limited by a radius of about four miles . . . a horse-using city may grow out to seven or eight. . . . Is it too much . . . to expect that the available area for even the common daily toilers of the great city of year 2000 . . . will have a radius of over one hundred miles?' (pp. 51–2).

'What will be the forces acting upon the prosperous household . . . ? The passion for nature . . . the allied charm of cultivation . . . [and] that craving . . . for a little private *imperium* [are] the chief centrifugal inducements' (pp. 54–6). 'The only reason why they should get out to a residence here rather than there is the necessity of saving time, and such a violent upward gradient of fares as will quite out-balance the downward gradient of ground values' (p. 61). 'The general nature of the expansion of the great cities in the future . . . [therefore] will not be a regular diffusion . . . but a process of throwing out the "homes" and of segregating various types of people. The omens seem to point pretty unmistakably to a wide and quite unprecedented diversity in the various suburban . . . districts' (p. 62).

'The city will diffuse itself until it has taken up considerable areas and many of the characteristics of what is now country. . . . The country will take itself many of the qualities of the city. The old antithesis will . . . cease, the boundary lines will altogether disappear. . . . To receive the daily paper a few hours late . . . will be the extreme measure of rusticity save in a few remote islands and inaccessible places' (pp. 70–71). '"Town" and "city" will be, in truth, terms as obsolete as "mail coach". . . . We may call . . . these coming town provinces "urban regions"' (p. 67).

#### URBAN REGIONS AND INTER-URBAN PERIPHERIES OF THE UNITED STATES IN 1960

Achieved at a pace more rapid than he anticipated, Wells's forecasts captured the essential features of the geography of the United States in 1960. 'City' and the continuously built-up 'urbanized area' had already been superseded in the realities of daily life by larger urban regions—commuting areas or, as C. A. Doxiadis (1969) calls them, 'daily urban systems'.

The daily urban system of Chicago, to look at one example, had a radius of up to 160 km in 1960 (Fig. 1). The dominant commuting flow was still to and from more highly concentrated and centralized workplaces, so that the intensity of long-distance commuting

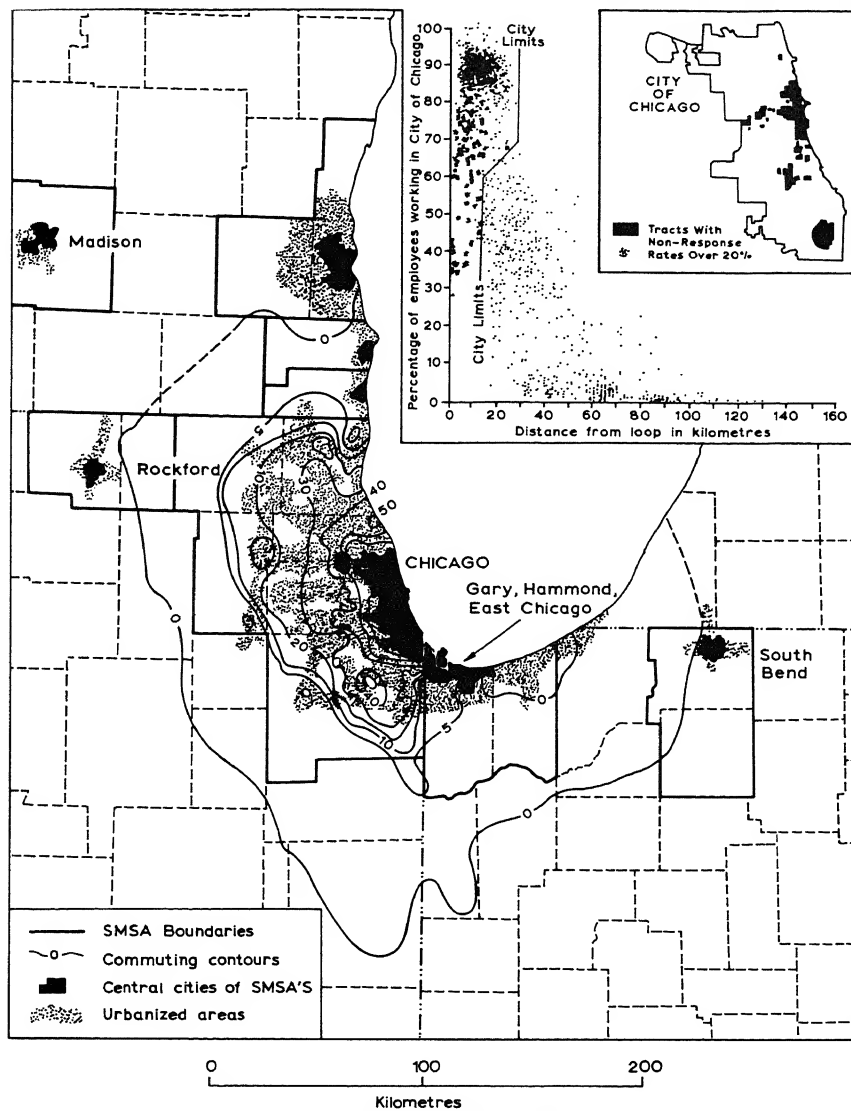


FIGURE 1. The commuting field of Chicago

declined with distance from the central city (as revealed in the inset graph on Figure 1, which plots every census tract in the daily urban system). But the local diversity was very great and the distance-decay has been decreasing progressively, typical of a highly differen-

tiated social geography and a rapidly decentralizing industry (Berry and F. Horton, 1970). Moreover, a substantial zone of the central city showed an exceptionally low degree of participation in the labour market. This zone was the reception area for migrants from rural poverty areas.

In 1960, all but 5 per cent of the nation's population lived within such daily urban systems, which sprawled like amoebae over the landscape (Fig. 2). If one calls the zones beyond the daily urban systems the 'inter-urban periphery', and then traces out combinations of commuter-sheds and boundaries between television viewing areas, to add a further fundamental element of daily interaction, a fairly distinct regionalization of the country emerges (Fig. 3).<sup>1</sup> For the purist who likes an exclusive and exhaustive regionalization, addition of the wholesale trade relationships of the daily urban systems permits a complete disaggregation of the conterminous United States into 171 economic areas, each relatively self-sufficient in the tertiary sector (Fig. 4).<sup>2</sup>

Within each region, regular gradients of metropolitan influence can be traced, revealing that identical rhythms characterize a broad range of social and economic phenomena (Fig. 5). As the degree of commuting declines, so do median incomes, but the proportion of the population below the poverty line increases. The peripheral zones of poverty are the zones of emigration to the city centres, helping to explain the pockets of poverty accumulating there, and both city centre and inter-urban periphery are, in general, zones of population decline.

Juxtaposing the commuting areas of the nation's seventy-odd truly metropolitan centres and the rest of the country, we therefore find that:

1. Most parts of the nation lying outside these areas are experiencing population decline (Fig. 6).
2. This is largely because of migration out of the inter-urban peripheries (Fig. 7) into the major metropolitan areas (Figs 8 and 9).<sup>3</sup>
3. The only exceptions are certain peripheral areas of high fertility, associated with particular minority groups (Fig. 10).
4. The result is two kinds of periphery, that with a declining, elderly white population, and that occupied by minority groups in which natural increase is able to feed emigration and still maintain a youthful population (Fig. 11).
5. All minority groups are now affected. Whereas a decade ago the American Indian was still restricted to the reservations to which he had been forced by Federal directive (Fig. 12), in the last decade he has urbanized apace (Fig. 13; E. Neils, 1969).

#### EMERGENCE OF THE PATTERN

Most American scholars today agree that the broad processes of national development that have, together, resulted in this geography of the United States as it appeared in 1960 have been urban-centred from the start. To be sure, '... throughout the evolution... two factors, great migrations and major changes in technology, have particularly influenced the location of relative growth and decline. ... Major changes in technology have resulted in critically important changes in the evaluation or definition of particular resources on which the growth of certain urban regions had previously been based. Great migrations have sought to exploit resources—ranging from climate or coal to water to zinc—that were newly appreciated or newly accessible within the national market. Usually... the new

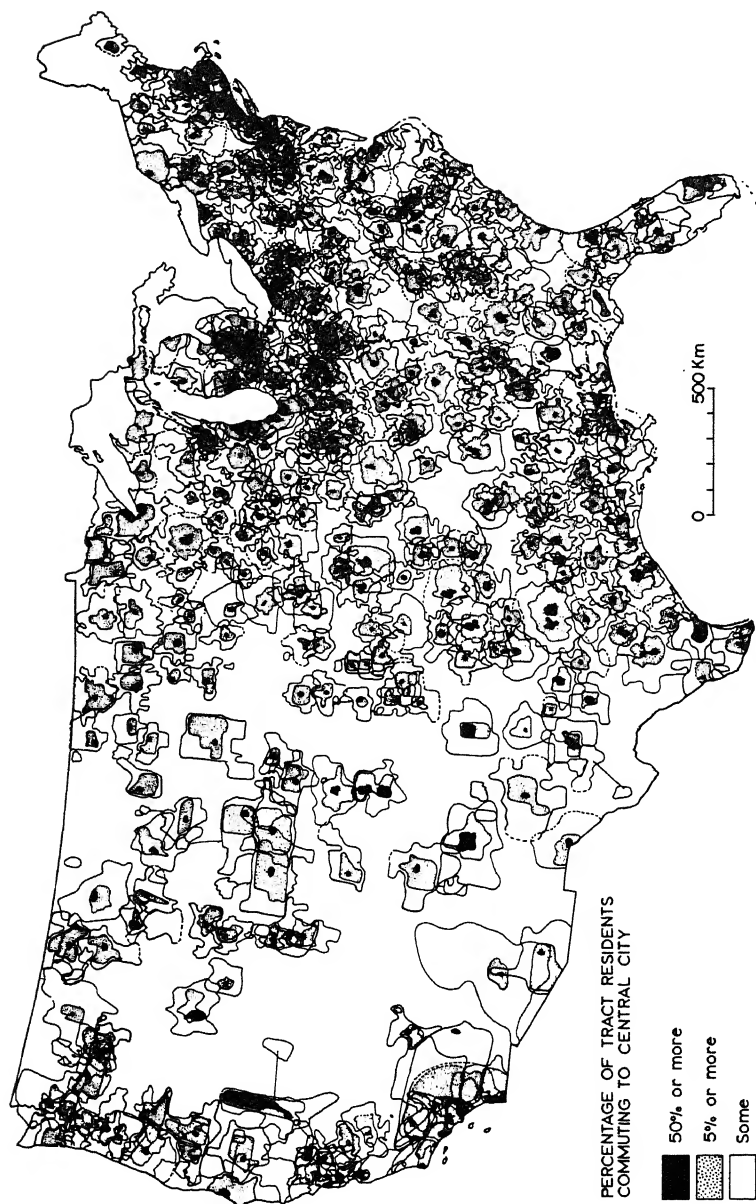


FIGURE 2. Commuting fields of all major central cities in the United States

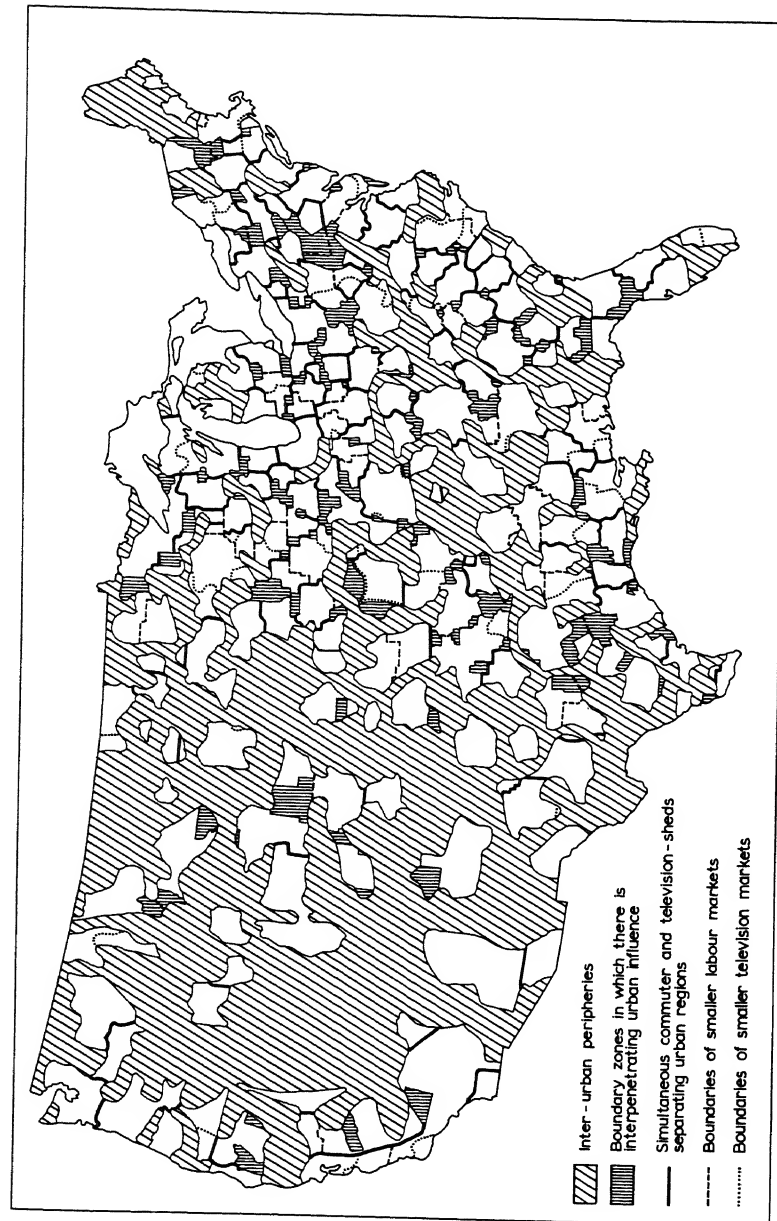


FIGURE 3. Daily urban systems of the United States

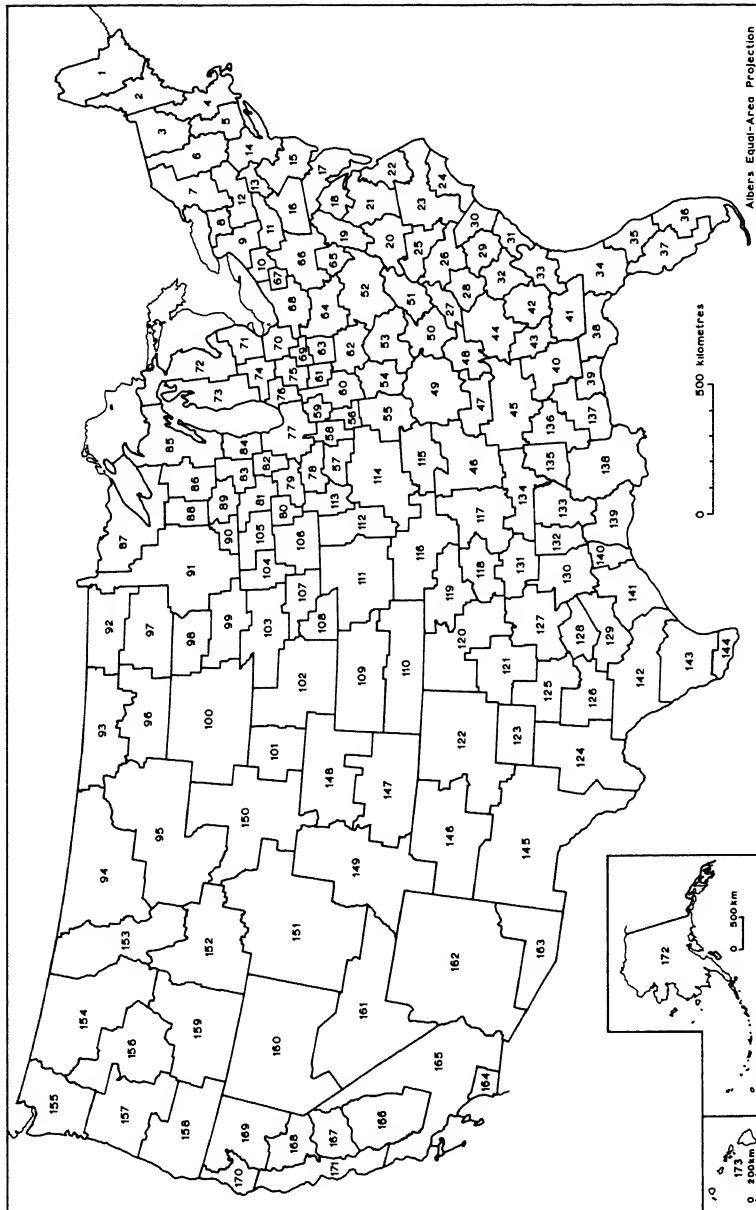
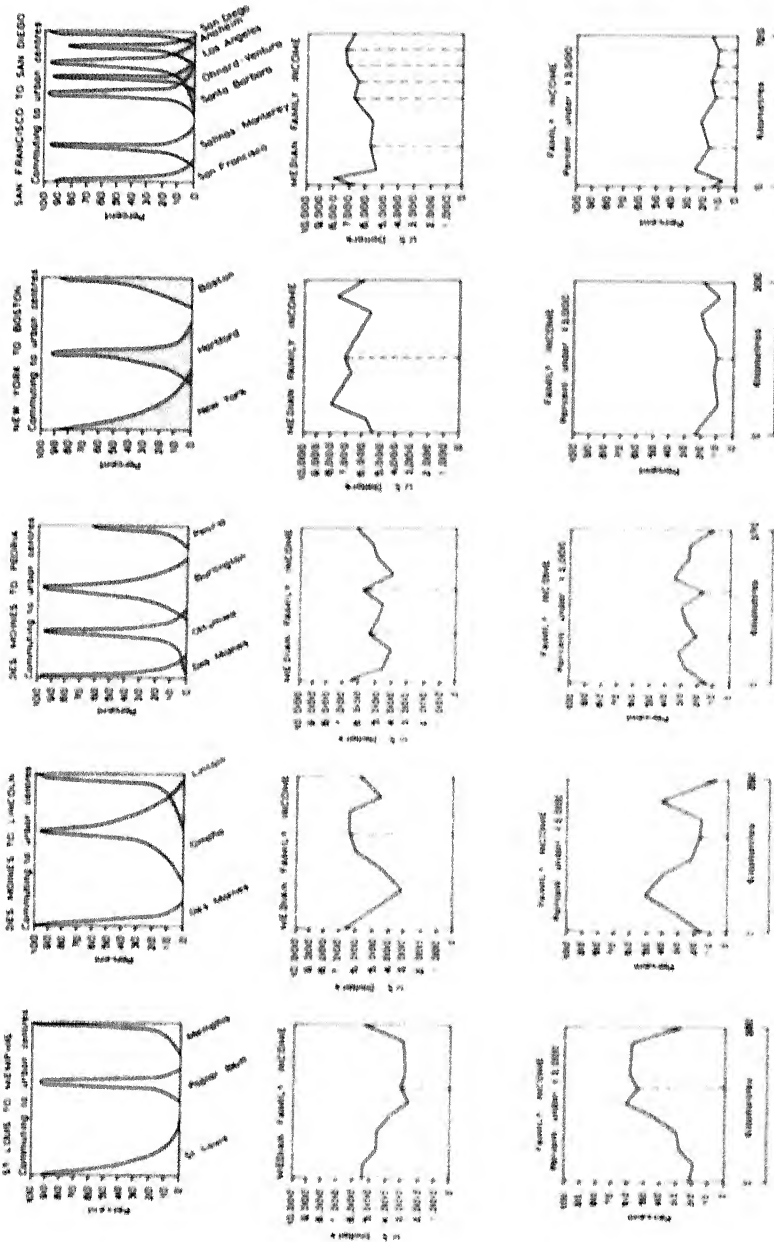


FIGURE 4. Functional economic areas maximizing closure in the residential sector

1. Bangor, Maine
2. Portland, Maine
3. Burlington, Vt.
4. Boston, Mass.
5. Hartford, Conn.
6. Albany-Schenectady-Troy, N.Y.
7. Syracuse, N.Y.
8. Rochester, N.Y.
9. Buffalo, N.Y.
10. Erie, Pa.
11. Williamsport, Pa.
12. Binghamton, N.Y.-Pa.
13. Wilkes-Barre-Hazleton, Pa.
14. New York, N.Y.
15. Philadelphia, Pa.-N.J.
16. Harrisburg, Pa.
17. Baltimore, Md.
18. Washington, D.C.-Md.-Va.
19. Staunton, Va.
20. Roanoke, Va.
21. Richmond, Va.
22. Norfolk-Portsmouth, Va.
23. Raleigh, N.C.
24. Wilmington, N.C.
25. Greensboro-Winston Salem-High Point, N.C.
26. Charlotte, N.C.
27. Asheville, N.C.
28. Greenville, S.C.
29. Columbia, S.C.
30. Florence, S.C.
31. Charleston, S.C.
32. Augusta, Ga.
33. Savannah, Ga.
34. Jacksonville, Fla.
35. Orlando, Fla.
36. Miami, Fla.
37. Tampa-St. Petersburg, Fla.
38. Tallahassee, Fla.
39. Pensacola, Fla.
40. Montgomery, Ala.
41. Albany, Ga.
42. Macon, Ga.
43. Columbus, Ga.-Ala.
44. Atlanta, Ga.
45. Birmingham, Ala.
46. Memphis, Tenn.-Ark.
47. Huntsville, Ala.
48. Chattanooga, Tenn.-Ga.
49. Nashville, Tenn.
50. Knoxville, Tenn.
51. Bristol, Va.-Tenn.
52. Huntington-Ashland, W.Va.-Ky.-Ohio
53. Lexington, Ky.
54. Louisville, Ky.-Ind.
55. Evansville, Ind.
56. Terre Haute, Ind.
57. Springfield, Ill.
58. Champaign-Urbana, Ill.
59. Lafayette-West Lafayette, Ind.
60. Indianapolis, Ind.
61. Muncie, Ind.
62. Cincinnati, Ohio-Ky.-Ind.
63. Dayton, Ohio
64. Columbus, Ohio
65. Clarksburg, W.Va.
66. Pittsburgh, Pa.
67. Youngstown-Warren, Ohio
68. Cleveland, Ohio
69. Lima, Ohio
70. Toledo, Ohio
71. Detroit, Mich.
72. Saginaw, Mich.
73. Grand Rapids, Mich.
74. Lansing, Mich.
75. Fort Wayne, Ind.
76. South Bend, Ind.
77. Chicago, Ill.
78. Peoria, Ill.
79. Davenport-Rock Island-Moline, Iowa -Ill.
80. Cedar Rapids, Iowa
81. Dubuque, Iowa
82. Rockford, Ill.
83. Madison, Wis.
84. Milwaukee, Wis.
85. Green Bay, Wis.
86. Wausau, Wis.
87. Duluth-Superior, Minn.-Wis.
88. Eau Claire, Wis.
89. La Crosse, Wis.
90. Rochester, Minn.
91. Minneapolis-St. Paul, Minn.
92. Grand Forks, N.D.
93. Minot, N.D.
94. Great Falls, Mont.
95. Billings, Mont.
96. Bismarck, N.D.
97. Fargo-Moorhead, N.D.-Minn.
98. Aberdeen, S.D.
99. Sioux Falls, S.D.
100. Rapid City, S.D.
101. Scotts Bluff, Nebr.
102. Grand Island, Nebr.
103. Sioux City, Iowa-Nebr.
104. Fort Dodge, Iowa
105. Waterloo, Iowa
106. Des Moines, Iowa
107. Omaha, Nebr.-Iowa
108. Lincoln, Nebr.
109. Salina, Kans.
110. Wichita, Kans.
111. Kansas City, Mo.-Kans.
112. Columbia, Mo.
113. Quincy, Ill.
114. St. Louis, Mo.-Ill.
115. Paducah, Ky.
116. Springfield, Mo.
117. Little Rock-North Little Rock Ark.
118. Fort Smith, Ark.-Okla.
119. Tulsa, Okla.
120. Oklahoma City, Okla.
121. Wichita Falls, Tex.
122. Amarillo, Tex.
123. Lubbock, Tex.
124. Odessa, Tex.
125. Abilene, Tex.
126. San Angelo, Tex.
127. Dallas, Tex.
128. Waco, Tex.
129. Austin, Tex.
130. Tyler, Tex.
131. Texarkana, Tex.-Ark.
132. Shreveport, La.
133. Monroe, La.
134. Greenville, Miss.
135. Jackson, Miss.
136. Meridian, Miss.
137. Mobile, Ala.
138. New Orleans, La.
139. Lake Charles, La.
140. Beaumont-Port Arthur-Orange, Tex.
141. Houston, Tex.
142. San Antonio, Tex.
143. Corpus Christi, Tex.
144. Brownsville-Harlingen-San Benito, Tex.
145. El Paso, Tex.
146. Albuquerque, N.M.
147. Pueblo, Col.
148. Denver, Col.
149. Grand Junction, Col.
150. Cheyenne, Wyo.
151. Salt Lake City, Utah
152. Idaho Falls, Idaho
153. Butte, Mont.
154. Spokane, Wash.
155. Seattle-Everett, Wash.
156. Yakima, Wash.
157. Portland, Ore.-Wash.
158. Eugene, Ore.
159. Boise City, Idaho
160. Reno, Nev.
161. Las Vegas, Nev.
162. Phoenix, Ariz.
163. Tucson, Ariz.
164. San Diego, Calif.
165. Los Angeles-Long Beach, Calif.
166. Fresno, Calif.
167. Stockton, Calif.
168. Sacramento, Calif.
169. Redding, Calif.
170. Eureka, Calif.
171. San Francisco-Oakland, Calif.
172. Anchorage, Alaska
173. Honolulu, Hawaii





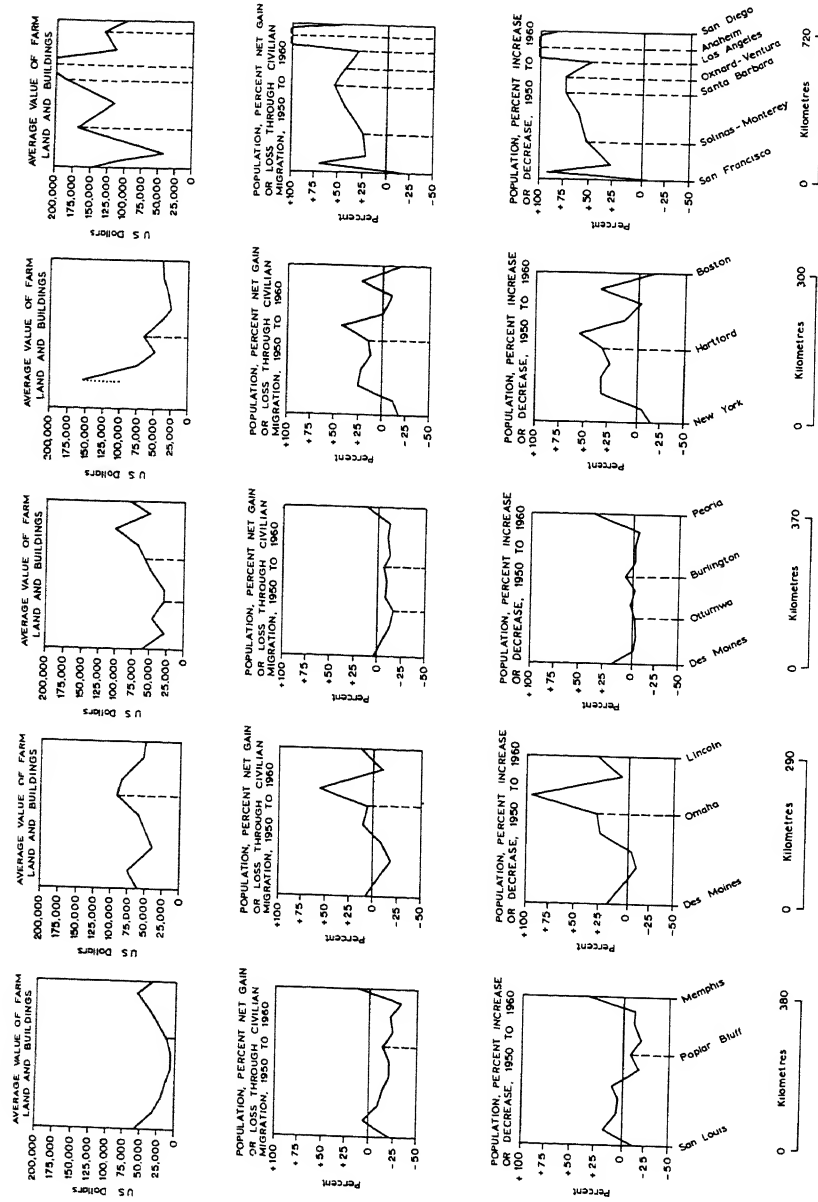


FIGURE 5. Gradients of urban influence

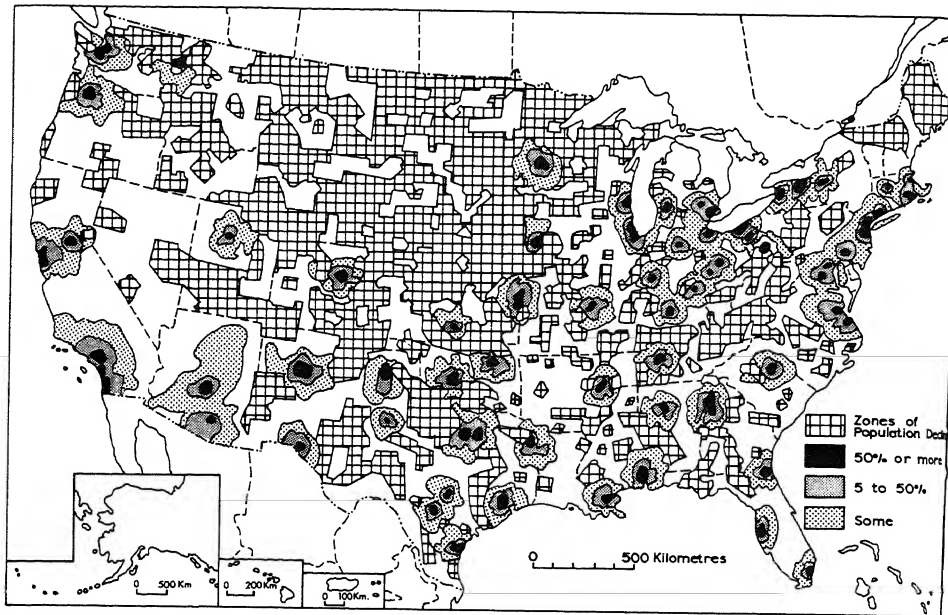


FIGURE 6. Zones of population decline in the United States, 1950-60

appreciation or accessibility has come about, in turn, through some major technological innovation' (J. R. Borchert, 1967, p. 324).

However, each of the successive sets of 'natural resources that count' has had an urban-centred definition. In the agricultural period, the natural endowment most valued was arable land with environmental components of climate and water. During industrialization, a new set of mineral resources became important. In the twentieth century, service activities and amenity resources have exerted an increasingly strong pull on industry and on people. But in each of these stages, urban growth and regional development have been closely interdependent.

North America's oldest cities were mercantile outposts of a resource area whose exploitation was organized by the developing metropolitan system of western Europe. The initial impulses for independent urban growth came at the end of the eighteenth century, when towns were becoming the outlets for capital accumulated in commercial agriculture and the centres of colonial development of the continental interior. Regional economies developed a certain archetype: a good deep-water port as the nucleus of an agricultural hinterland well adapted for the production of a staple commodity in demand on the world market. Growth potentials of regions depended on the extent and richness of the hinterlands accessible to the ports. The prototype of the American metropolis thus was a port at a strategic location on long-distance oceanic or riverine trade routes, providing a range of mercantile services, and determining the terms of trade. The growth of this agricultural

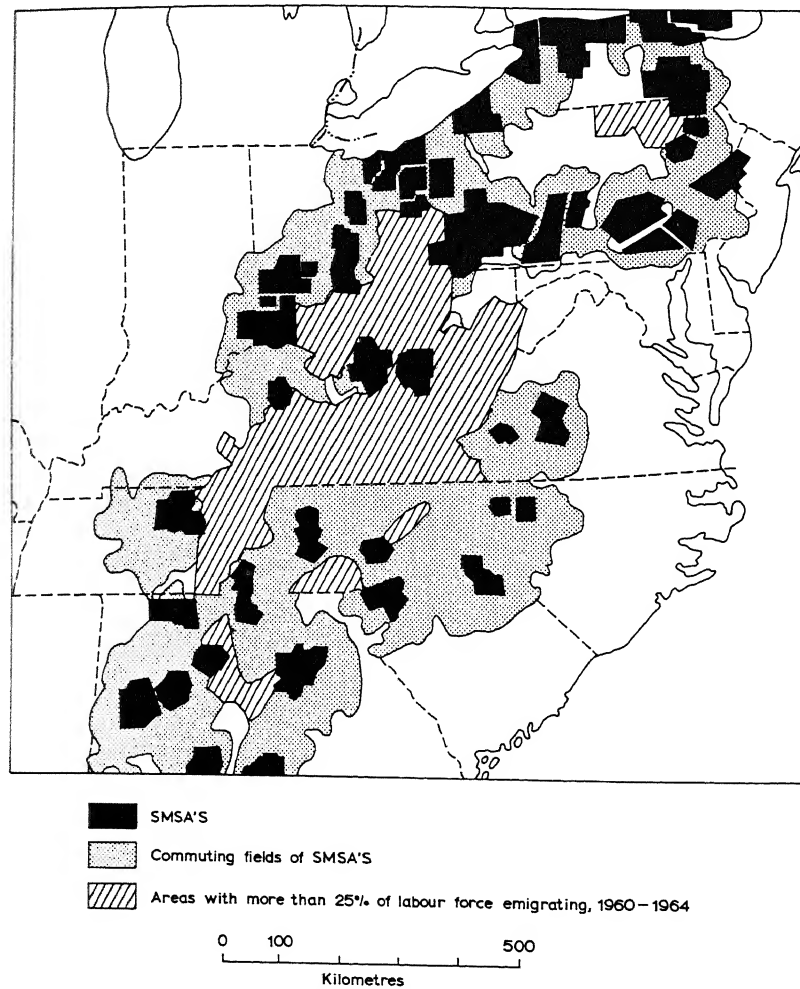


FIGURE 7. Areas with high emigration rates within Appalachia, 1960-64

resource-dominated (but city-centred) expansion of the economy set the stage for subsequent developments by establishing a geography of markets, transport routes and labour forces that conditioned the nature of succeeding growth.

New resources became important from 1840-50 onward, and new locational forces came into play. Foremost was a growing demand for iron, and later steel, and along with it rapid elaboration of productive technologies. Juxtaposition of coal, iron ore and markets

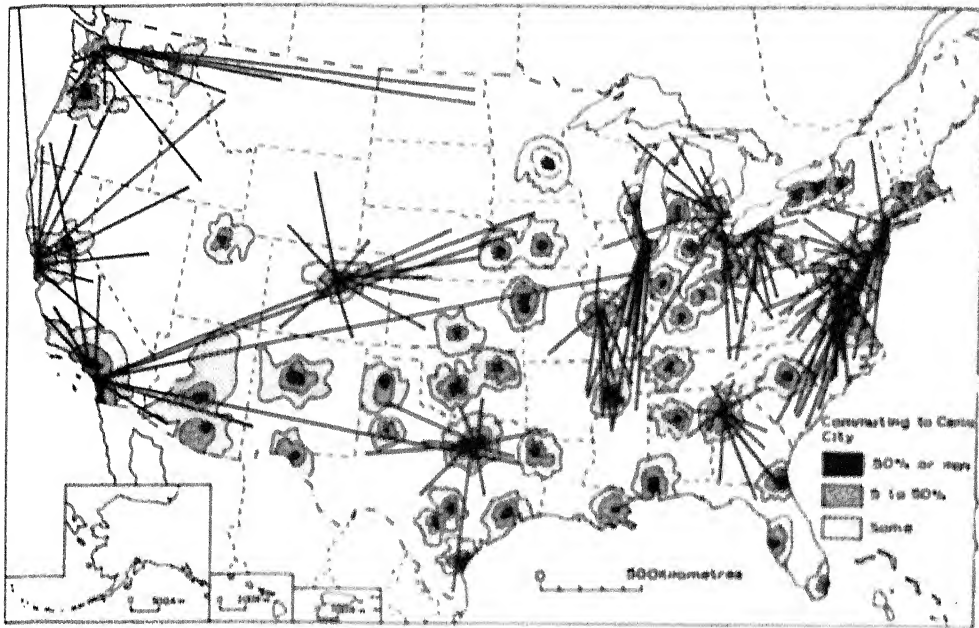


FIGURE 8 Migration fields of the fourteen largest SMSAs, 1955-60

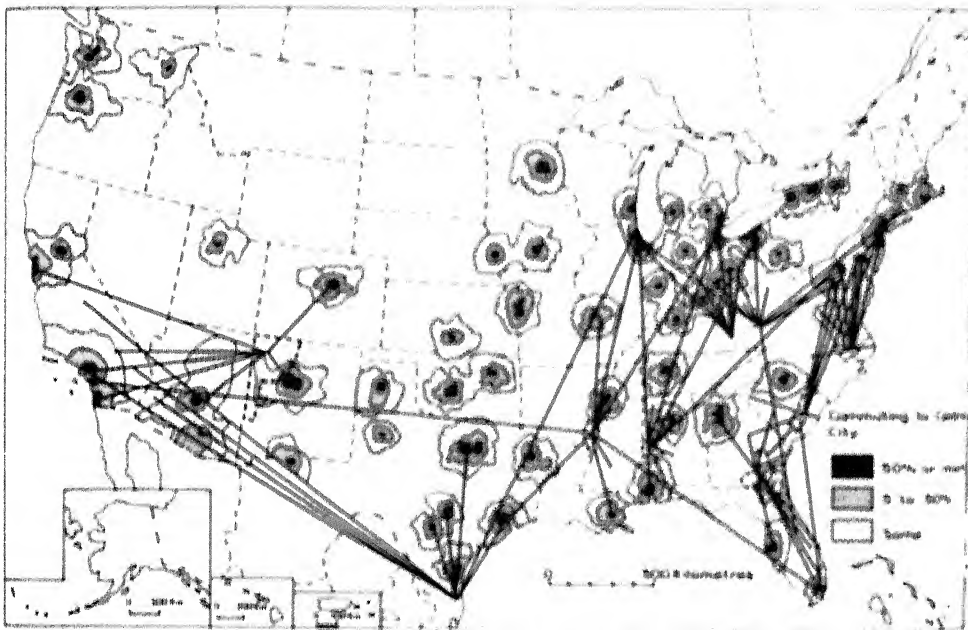


FIGURE 9 Principal destinations of migrants from selected poverty regions, 1955 for SMSAs for these and the next three Figures are as defined in Figure 8

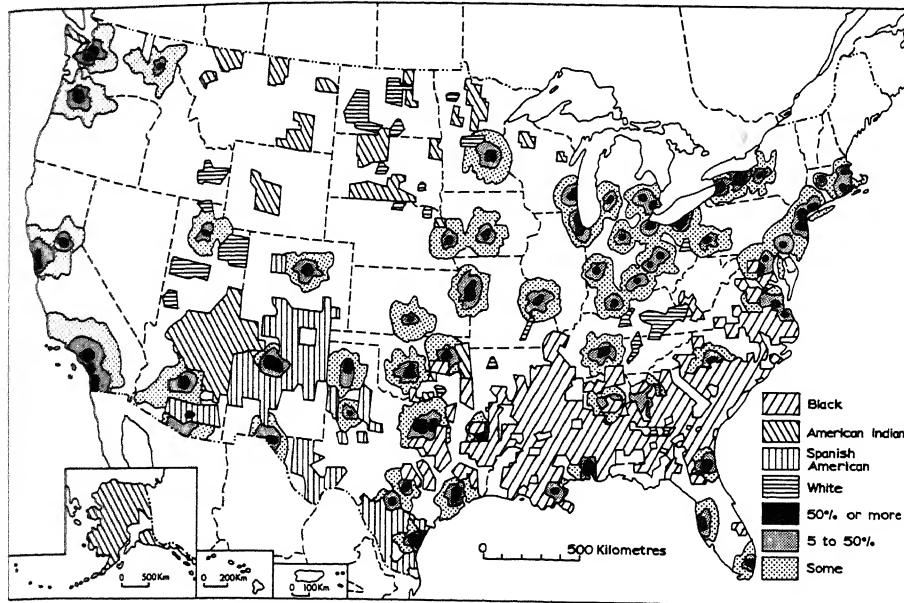


FIGURE 10. Areas of high fertility in 1960, classified by race

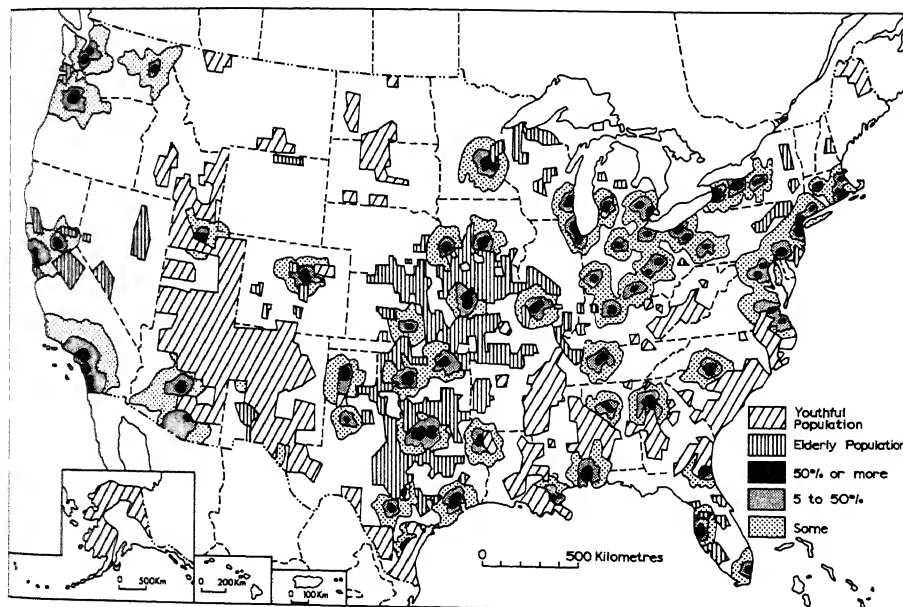


FIGURE 11. Areas of youthful and elderly population in 1960

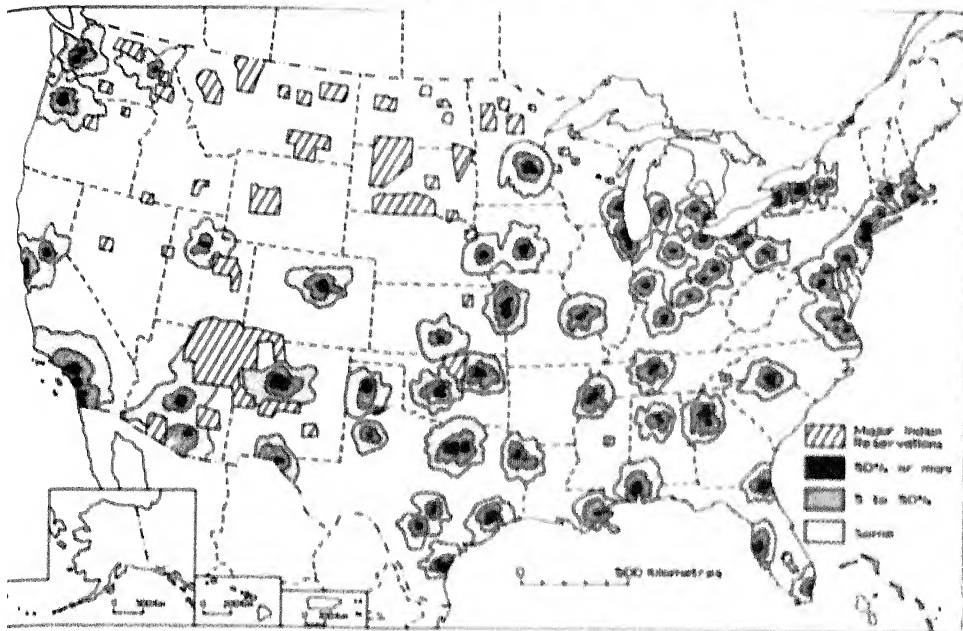


FIGURE 14 Major Indian reservations

afforded the impetus for manufacturing growth in the north-eastern United States, localized both by factors in the physical environment (minerals) and by environmental components created by prior growth of the urban system (linkages to succeeding stages of production, in turn located closer to markets). The 'heartland' of the North American manufacturing belt developed westward from New York in the area bounded by Lake Superior iron ores, the Pennsylvanian coalfields, and the capital, entrepreneurial experience and engineering trades of the north-east, while at the same time New York cemented its dominance by accentuation of its financial, entrepreneurial and specialized manufacturing roles (Fig. 14). This heartland became not only the heavy industrial centre of the country, but has remained the centre of national demand, determining patterns of market accessibility ever since (Fig. 15).

The heartland had initial advantages of both excellent agricultural resources and a key location in the minerals economy. With development, it grew into the urbanized centre of the national market. Subsequent metropolitan growth was in a pattern organized around this national core region until the last decade. From 1869 to the mid-1950s, there had been a stable pattern of growth in manufacturing employment among the states (G. H. Horts, 1967). Continued spread of population and agriculture over the continent pulled processing and servicing activities and new urban growth with them. However, the dominant effects in terms of development still came from growth of the minerals economy until well into the twentieth century, so that a process of 'circular and cumulative causation' (A. Pred, 1966) strengthened and maintained the relations of the national heartland and hinterlands of



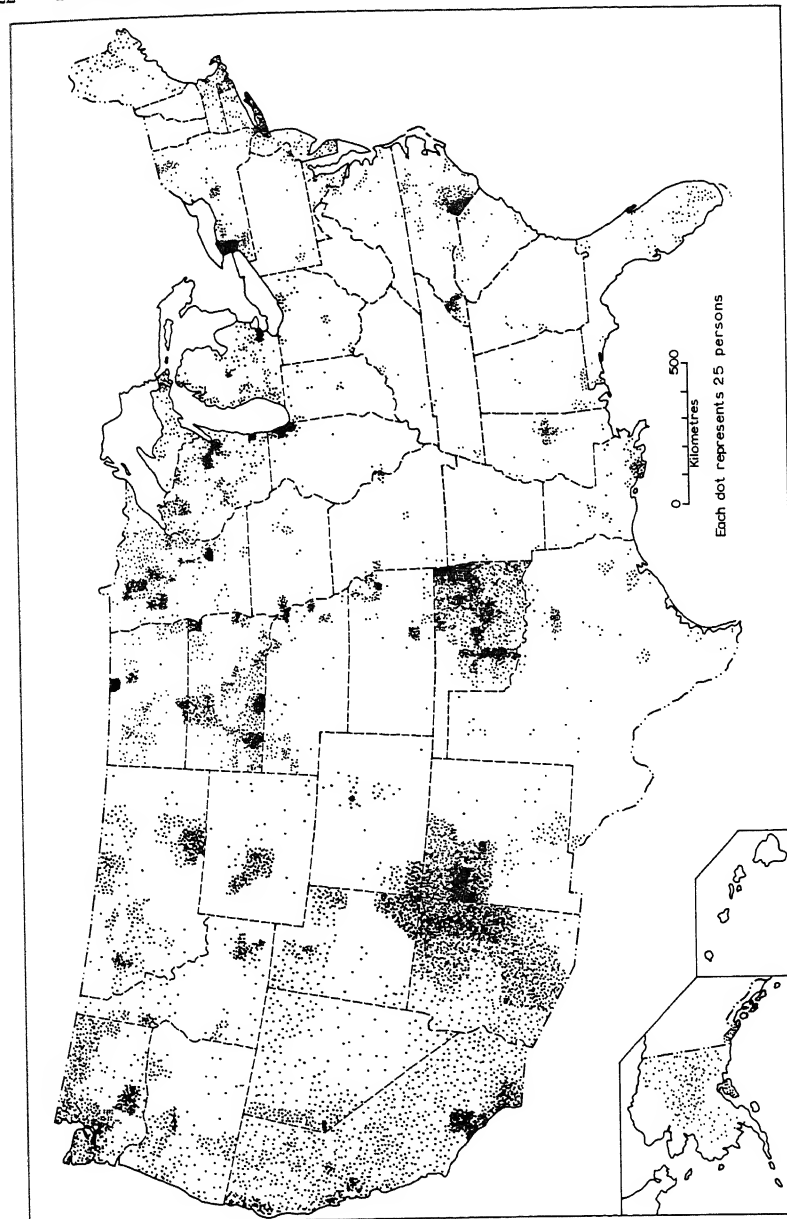


FIGURE 13. Distribution of the Indian population by county in 1960

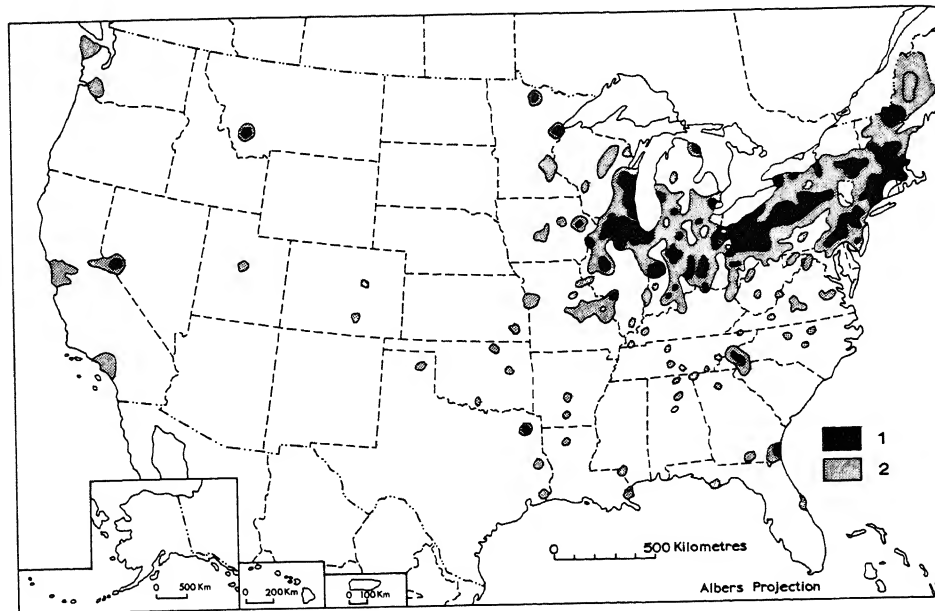


FIGURE 14. Areas with high dependence upon manufacturing in 1962. 1—Factor scores exceed  $+2.0$  in analysis of county employment structure; 2—Factor scores between  $+1.0$  and  $+1.99$

core and periphery—and the new metropolitan centres that did emerge did so in sequence with the overall growth, outward spread, and spatial integration of the economy.

In each case, the basic conditions of regional growth were set by the heartland. It was the lever for successive development of newer peripheral regions by reaching out to them as its input requirements expanded, and it thereby fostered specialization of regional roles in the national economy. The heartland experienced cumulative urban-industrial specialization, while each of the hinterlands found its comparative advantage based on narrow and intensive specialization in a few resource sub-sectors, only diversifying when the extent of specialization enabled the hinterland region to pass through that threshold scale of market necessary to support profitable local enterprise. Flows of raw materials inward, and of finished products outward, articulated the whole (E. Ullman, 1957).

It is little wonder that, in his masterful analysis of regional growth in the United States through the 1950s, H. S. Perloff (1960) concluded that the central driving force is '... a great heartland nucleation of industry and the national market, the focus of large-scale national-serving industry, the seedbed of new industry responding to the dynamic structure of national final demand and the center of high levels of per capita income...' and that, standing in a dependent relationship to the heartland, '... radiating out across the national landscape are ... resource-dominant regional hinterlands specializing in the production of resource and intermediate outputs for which the heartland reaches out to satisfy the input requirements of its great manufacturing plants. Here in the hinterlands, re-



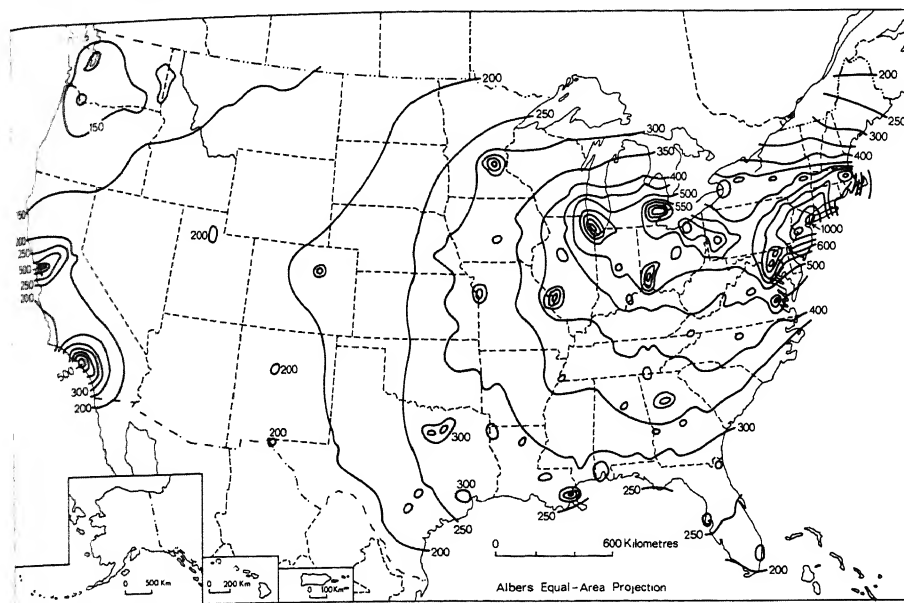


FIGURE 15. National market access as indexed by population potentials in 1960. Isolines represent population potentials in thousands per mile

source-endowment is a critical determinant of the particular cumulative advantage of the region and hence its growth potential.<sup>7</sup>

From this situation, of course, we derive our conventional current theory of urban and regional growth, which states that each region is an integral part of a larger interdependent economy, and so long as demand and supply conditions change and regions have differing advantages for production, differences in regional growth must be seen in larger systemic terms. Basic to the growth of given regions, according to the theory, is their capacity for attracting industries that produce goods for export to other regions, so that the nature of regional specialities and changes in the structure of demand for them determine in large measure the nature and extent of regional growth. In turn, the level of export activity determines the level of local market provision of goods and services for 'residential' demands.

Throughout, of course, the theory points out how important it is to remember the role of the metropolitan centres. It is they that organize the space-economy. They are the centres of activity and of innovation, focal points of the transport and communications networks, locations of superior accessibility at which firms can most easily reap scale economies and at which industrial complexes can obtain the economies of localization and urbanization. It is they, therefore, that encourage labour specialization, areal specialization in productive activities, and efficiency in the provision of services. Metropolitan centres, standing at the top of the nation's urban hierarchy and at the centre of each of its regions, are therefore the nodes whose connections integrate the national economy. Continuing

agricultural enterprise is more efficient in the vicinity of cities. The more prosperous commercialized agricultures encircle the major cities, whereas the peripheries of the great urban regions are characterized by backward lower-income economic systems.

The theory, derived as it is from studying the last century's economic history, goes on to say that, like regions, cities have similar national bases for their growth. The polarization of growth in cities, particularly larger metropolitan areas, is a reflection of trends in the national space economy. Differential growth rates among cities are a function of competitive advantages in location, accessibility and local economic mix. The characteristic structure of a system of cities at any point in time reflects past growth performance and provides the basic infrastructure for future growth.

But all of a sudden our urban economists have started to sing another tune. For example, W. Thompson (1968) has pointed out that, today, the long-run viability of a metropolitan

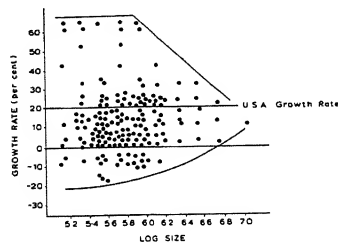


FIGURE 16. Growth rates of the daily urban systems, 1950-60, compared with their size in 1960

area resides not so much in regional resources, traditionally viewed, but on its own capacity to invent and innovate, or otherwise acquire new resources and export bases. To follow his argument, he says that the economic base of the larger metropolis is the creativity of its universities and research parks, the sophistication of its engineering firms and financial institutions, the persuasiveness of its public relations and advertising agencies, the flexibility of its transport networks and utility systems, and all the dimensions of 'infra-structure' that facilitate the quick and orderly transfer from old dying bases to new growing ones. Larger urban areas, he argues, combine a favourable mix of industry for growth with a steadily declining share of the various growth industries. High wage rates of the innovating area, quite consonant with the high skills needed at the beginning of the learning process, become excessive when skill requirements decline, and the industry (or parts of it) will then 'filter down' to smaller, less industrially sophisticated areas where the cheaper labour can meet the declining demands for skills of the filtering industry, thus creating the phenomenon of small towns with 'low-wage, slow-growth filtered-down industry' at the time when the metropolis has moved on to new bases. *This capacity of the large metropolis to invent new economic bases means that large urban regions have freed themselves of narrow export dependency, traditionally defined.* Further, the larger the metropolis, the freer it is from traditional restraints. At a sufficiently large scale, infrastructure and residential development make the growth of the large metropolitan complex essentially self-generative, with a tendency to grow at or about the rate of the nation (Thompson, *ibid.*). Figure 16 shows this convergence in the growth rates of the 171 daily urban systems with increasing size in the decade 1950-60.

This trend in the urban system has been accentuated since 1950 by the growth of the service sector, an increase in the number of 'footloose' industries (including final processing of consumer goods using manufactured parts, and the aircraft, aerospace and defence industries), rapid emergence of a 'quaternary' sector of the economy (involving, for example, the research and development industry), the expansion and inter-regional migration of the non-job-oriented population (for example, of retired people to Florida, Arizona and

California), and the overall rises in real incomes. These factors have all served, additionally, to produce yet another transformation of the economy and the urban system based upon new amenity resources (Perloff, 1960). Advantages for economic growth have been found around the western and southern 'outer rim' of the country, in regions and places relatively well-endowed with such amenities, and in the outer rim of each urban field, as changing technology of communications has reduced the time and costs involved in previous heartland-hinterland relationships and in communications between the centre and periphery of each field.

The advantages for the outlying regions have been cumulative, for regional growth within the context of the national pattern of heartland and hinterland had brought these regions to threshold sizes for internal production of a wide variety of goods and services at the very time that changes in the definition of urban resources made possible a rapid advance based upon superior factor endowments of the periphery; the combination pushed them beyond the threshold for self-sustaining growth. Hence the explosive metropolitan development in the south, south-west and west since 1950, and the equally explosive realignments within urban fields surrounding the older metropolises elsewhere in the nation.

Now there are various ways in which one can interpret the result. I like the one proposed by J. Friedmann and J. Miller (1965), who say that it is possible today '... to interpret the spatial structure of the United States in ways that will emphasize a pattern consisting of *one*, metropolitan areas and *two*, the inter-metropolitan periphery. Except for thinly populated parts of the American interior, the inter-metropolitan periphery includes all the areas that intervene among metropolitan regions and that are, as it were, the reverse image of the trend towards large scale concentrated settlement that has persisted in this country for over half a century. Like a devil's mirror, much of the periphery has developed a socio-economic profile that perversely reflects the very opposite of metropolitan virility'. This was the geography of the United States as we presented it in 1960.

#### CONTRASTING VIEWS OF THE FUTURE

Friedmann and Miller rightly go on to point out that the periphery has a disproportionately large share of low growth and declining industries and a correspondingly antiquated economic structure—an area in long-term and continuous decline, with selective emigration that has polarized the remaining population around the very young and the very old. The quality of public services has deteriorated, the housing stock is older, and educational attainment is significantly below that for metropolitan America. 'Rapid and selective outmigration, a declining economic base, the burden of an ageing population, and low incomes have rendered many peripheral communities helpless in their desire to adapt to changing circumstances in the outside world. The remaining population is frequently short both on civic leadership and hope. They can neither grasp the scope of the events that have overtaken them nor are they capable of responding creatively to the new situations' (Friedmann and Miller, *ibid.*).

In trying to diagnose the reasons for the pattern, Friedmann and Miller say 'The emergence in large sections of the country of the inter-metropolitan periphery as a major problem area has been the direct result of the concentration of people and activities around closely contiguous metropolitan cores. Growth in and around these cores has drawn off the productive population, economic activities, and investment capital of the periphery.'

On the other hand, they say 'Looking ahead to the next generation, we foresee a new

scale of urban living that will extend far beyond existing metropolitan cores and penetrate deeply into the periphery. Relations of dominance and dependency will be transcended. The older established centers, together with the intermetropolitan peripheries that envelop them, will constitute the new ecological unit of America's post-industrial society that will replace traditional concepts of the city and metropolis. This basic element of the emerging spatial order we shall call the *urban field*.

'The urban field may be viewed as an enlargement of the space for urban living that extends far beyond the boundaries of existing metropolitan areas—defined primarily in terms of commuting to a central city of "metropolitan" size into the open landscape of the periphery. This change to a larger scale of urban life is already under way, encouraged by changes in technology, economics, and preferred social behaviour'.

Spelling out their idea further, '... an urban field is based on the criterion of interdependency. It represents a fusion of metropolitan spaces and nonmetropolitan peripheral spaces centered upon core areas of at least 300,000 people and extending outwards from these core areas for a distance equivalent to two hours' driving over modern throughway systems (approximately 100 miles with present technology). This represents not only an approximate geographic limit for commuting to a job, but also the limit of intensive weekend and seasonal use (by ground transportation) of the present periphery for recreation' (Friedmann and Miller, *ibid.*).

But this is a conservative forecast indeed, for all that the authors are saying is that what exists will be, only in slightly greater degree. By 1960, the effect of American growth had already been to lay down on the country a system of spatial organization and resource evaluation based on a network of metropolitan centres essentially self-generating in growth and therefore transcending the older national order of heartland and hinterlands, with radiating effects of metropolitan influence decaying from core to periphery within each urban field.

That distance-decay should continue to lessen is hardly surprising and, as it does, it should undoubtedly lead Friedmann and Miller's network of urban fields in the direction of H. Kahn and A. Wiener's propositions (1967) that 'the United States of the year 2000 will see at least three gargantuan metropolises, *Boswash*, *Chipitts*, and *Sansan*, which should contain more than half the U.S. population, including an overwhelming proportion of the most technologically and scientifically advanced, the most prosperous and creative elements of society'. Kahn and Wiener continue that, while all three will be recognizably American in culture, each will most likely have a quite distinguishable sub-culture—the continuing cosmopolitanism of *Boswash*; the combination of Bible belt and Carl Sandburg's 'raw and lusty vitality', in *Chipitts*; and the 'wholesome degeneracy' of the informal 'barbecue' culture of *Sansan*. Involving the emergence of post-industrial society, study of the United States in the year 2000, Kahn and Wiener conclude, may largely be of *Boswash*, *Chipitts* and *Sansan*. They are supported by projections such as those of J. P. Pickard (1967), based on continuation of present trends, and those of the Office of Business Economics using similar methods. Both forecast a population in excess of 60 million in both the Atlantic Seaboard and Great Lakes urban regions by 2000, and another 40 million or so in California.

#### GROWTH PROCESSES AND AN ALTERNATIVE VIEW

These forecasts are conservative. They err where H. G. Wells erred—in underestimating the pace of change. They miss what J. J. Servan-Schreiber (1969) has called the essence of 'The

American challenge', the compression of time and space in a way that was inconceivable even 10 years ago, with attendant intensification of human experience alongside lessening demands for movement because of centralized information sources and instantaneous communication. Servan-Schreiber points out that the time-lag between invention and manufacture was 112 years for photography (1727-1839), 56 years for the telephone (1820-76), 35 years for radio (1867-1902), 15 years for radar (1925-40), 12 years for television (1922-34), 5 years for the transistor (1948-53) and 3 years for the integrated circuit (1958-61). I agree with Servan-Schreiber that this convergence is the most salient characteristic of American life today. We are on the verge of yet another fundamental transformation of American society, one destined, I anticipate, to *invert* the spatial patterns of 1960 by the year 2000.

To understand the logic of this anticipation, we must first understand the nature of the growth processes currently operating in the country. There are two major elements in the city-centred organization of economic activities in space that are important:

- (a) *A system of cities*, arranged in a *hierarchy* according to the functions performed by each
- (b) Corresponding *areas of urban influence* surrounding each of the cities in the system.

We know the following about this system of spatial organization:

- (a) The size and functions of a central city, the size of its urban field, and the spatial extent of developmental 'spread effects' radiating outward from it are proportional.
- (b) Impulses of economic change are transmitted in order from higher to lower centres in the urban hierarchy, in a 'size-ratchet' sequence, so that continued innovation in large cities remains critical for extension of growth over the complete economic system.
- (c) The spatial incidence of economic growth is a function of distance from the central city. Troughs of economic backwardness lie in the most inaccessible areas along the peripheries between the least accessible lower-level centres in the hierarchy.
- (d) The growth potential of an area situated along an axis between two cities is a function of the intensity of interaction between them.

One would conclude from this that, if metropolitan development is sustained at high levels, differences between centre and periphery should be eliminated and the space-economy should be integrated by outward flows of growth impulses through the urban hierarchy, and the inward migration of labour to cities. Troughs of economic backwardness at the intermetropolitan periphery should, thereby, be eroded, and each area should find itself within the influence fields of a variety of urban centres of a variety of sizes. Continued

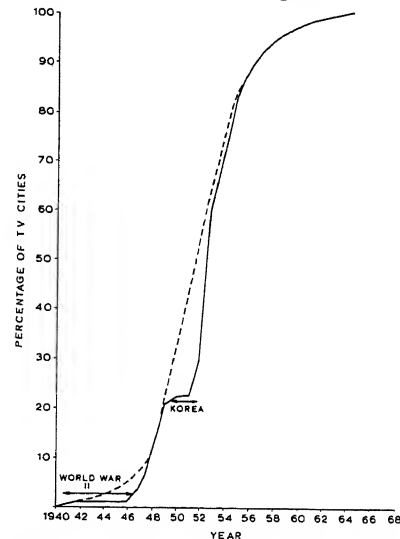


FIGURE 17. Growth in number of TV cities, 1940-68

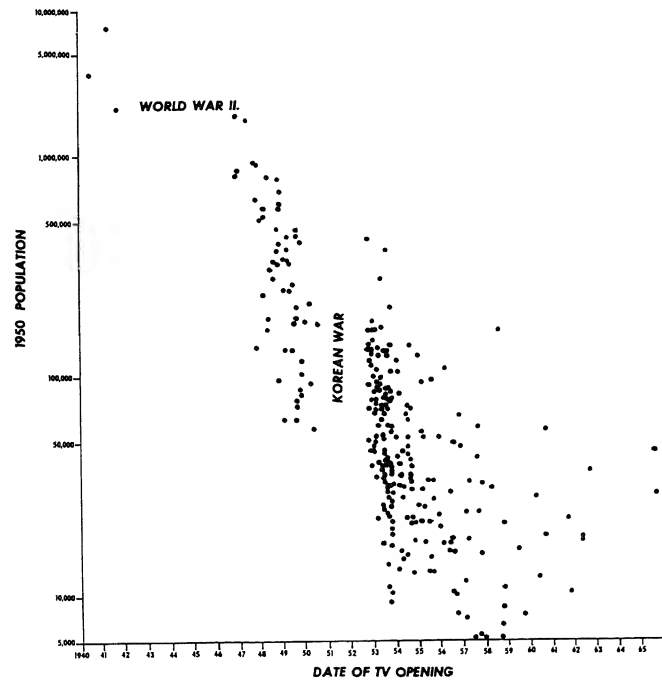


FIGURE 18. Hierarchical diffusion of the TV network

urban-industrial expansion in major central cities should lead to catalytic impacts on surrounding regions. Growth impulses and economic advancement should 'trickle down' to smaller places and ultimately infuse dynamism into even the most tradition-bound peripheries.

The recent diffusion of television illustrates these themes perfectly (Berry, 1970). Television stations were installed by cities in the familiar S-shaped logistic time sequence, although affected by two periods of war (Fig. 17). The diffusion pattern of stations among cities was essentially hierarchical. Large cities installed stations before smaller cities (Fig. 18). The proportion of households purchasing television sets depended upon the hierarchical diffusion of broadcasting stations, but the spatial incidence of purchases was also a function of distance from broadcasting cities (Fig. 19). As diffusion proceeded apace, market penetration neared completion first along the nation's principal growth axes and extended wavelike into the periphery (Fig. 20). Continuing, total geographical coverage was achieved by 1958 (Fig. 21), and low degrees of market penetration thereafter remained only in the nation's troughs of economic backwardness (Figs. 22 and 23).

The subtlety of television as an innovation is important, however, and this is why I chose it as an example of the traditional diffusion process. It arose as an 'evolutionary' innovation, one which is a logical extension of existing trends in the society and economy,

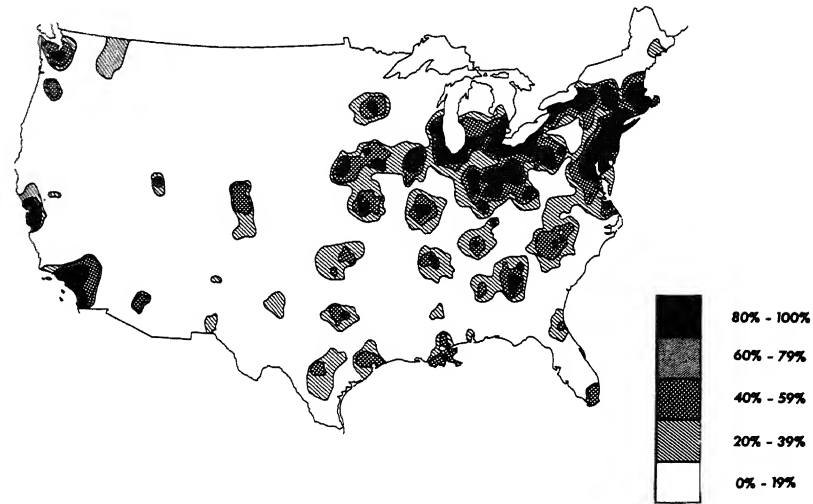


FIGURE 19. Market penetration by TV in 1953. Percentage figures on this and the next four Figures represent the percentage of households with TV.

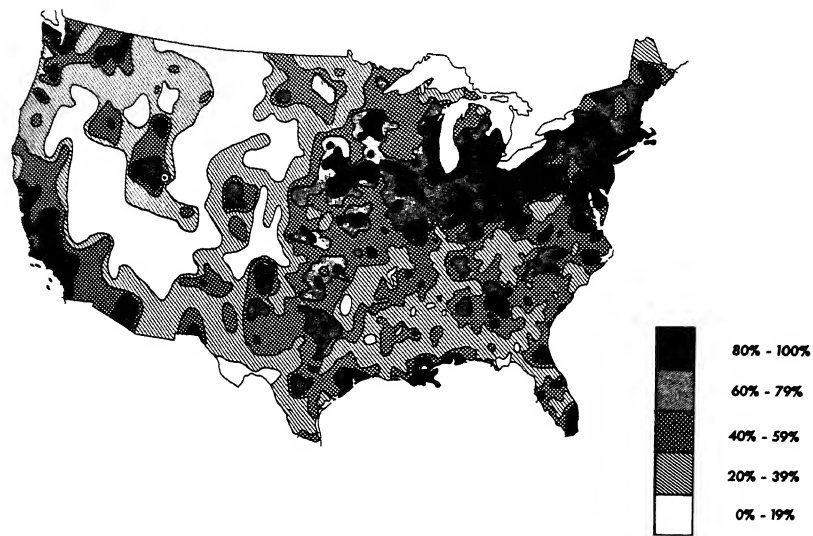


FIGURE 20. Market penetration by TV in 1956

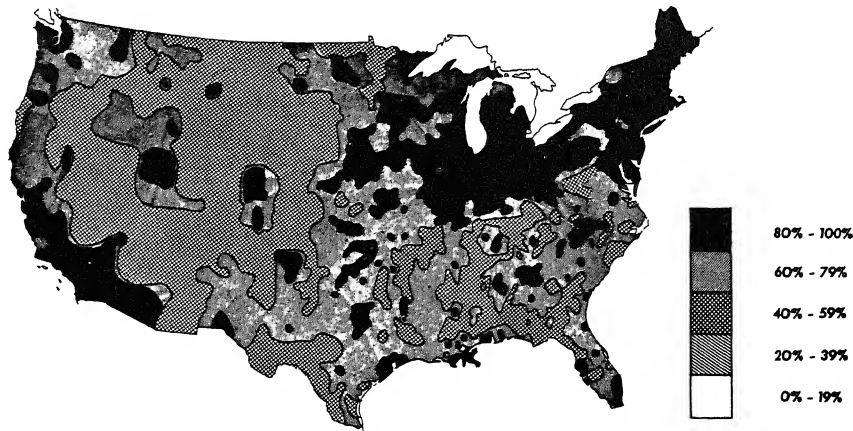


FIGURE 21. Market penetration by TV in 1959

and was adopted accordingly, both hierarchically and spatially, just as any other small adjustment or new technology has been in the United States. Yet it has had a revolutionary impact already, not simply because of its hypnotic domination of leisure time, but because of the universality and immediacy of experience that it brings into each household. The entirely new conditions thus created are not being countered by further incremental changes in the existing system, I submit, but by the emergence of entirely new social frameworks. Television, I think, is the first of a series of revolutionary electronic innovations that will affect America in the years to come.

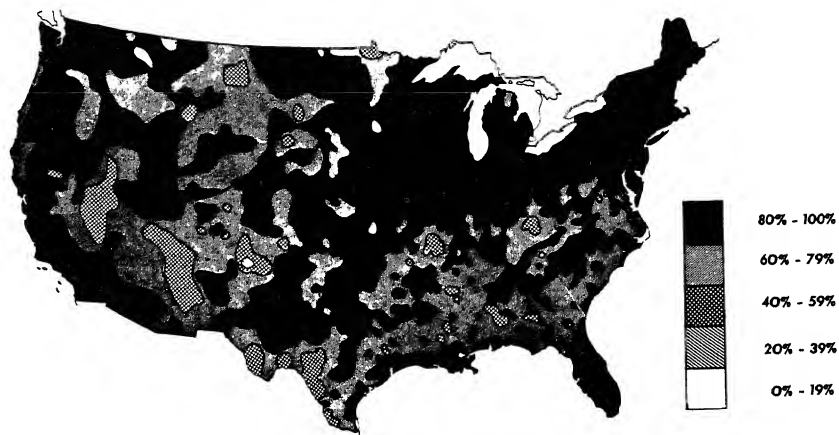


FIGURE 22. Market penetration by TV in 1962



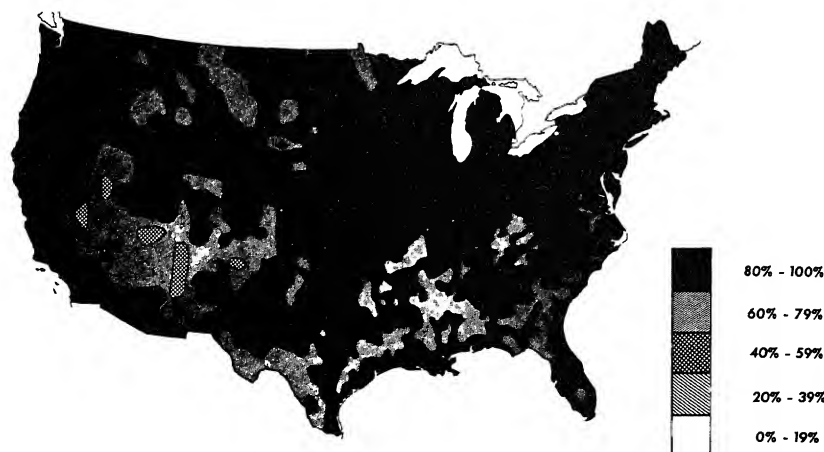


FIGURE 23. Market penetration by TV in 1965

The conditions are favourable. Rising real incomes, shortening working weeks, and increased leisure time all set the stage. So does the fact that most urban fields have now passed the minimum scale necessary for self-sustained growth either in simple economic terms or by use of political influence in Washington to guarantee continuing stability until they reach the size threshold—all of which means that the 'lever' effect of heartland growth upon hinterlands specializing in resource sub-sectors has vanished, or will soon do so, and along with it, alas, our traditional theories of urban and regional growth.

And now, as a result of reporting by the medium, we have the universal perception of decaying central cities, the new home of the former residents of the now-emptied periphery; the immediate on-the-spot experience of their riots; the careful documentation of their frustrations; and acute awareness of emerging separatist feelings. It is no accident that the suburbanization of white city dwellers has increased from the 450 000 net outflow annually of the 1950-60 decade to over 800 000 a year today.

All of this presages an inversion of American geography by the year 2000, an inversion already indicated in the contrasting pattern of acceptance of colour television in the years 1966-68 (Figs. 24, 25, 26). I foresee the legal central cities assuming a new reality as the bases within which the nation's poor minorities can gain and exert political power and influence. I also foresee that gradients of distance-accretion will replace those of distance-decay. Persons of greater wealth and leisure will find homes and work among the more remote environments of hills, water and forest, while most will aspire to this as an ideal. This, of course, is another inversion; the environments historically least valued are rapidly becoming those most desired. The signs are already there to be noted. As I wrote this paper, for example, I read of I. W. Abel, president of the United Steelworkers, calling for a four-day working week and a three-day weekend. And again, in *Time* magazine for 26 September 1969 about Windham County in south-eastern Vermont, a traditional Yankee area of woods, farms, and small towns, which is now being subdivided for second vacation

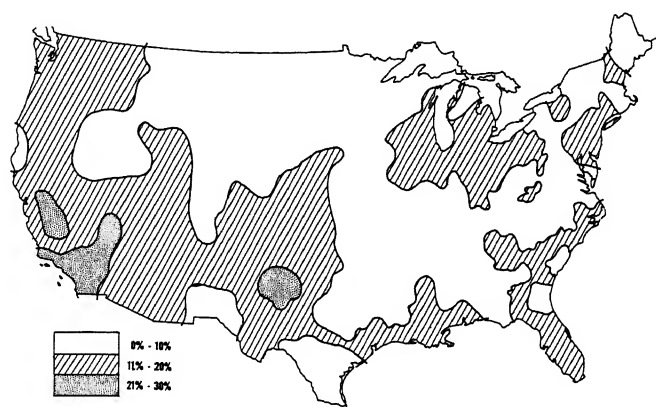


FIGURE 24. Market penetration by colour TV in 1966. Percentage figures on this and the next two Figures represent the percentage of TV households with colour TV.

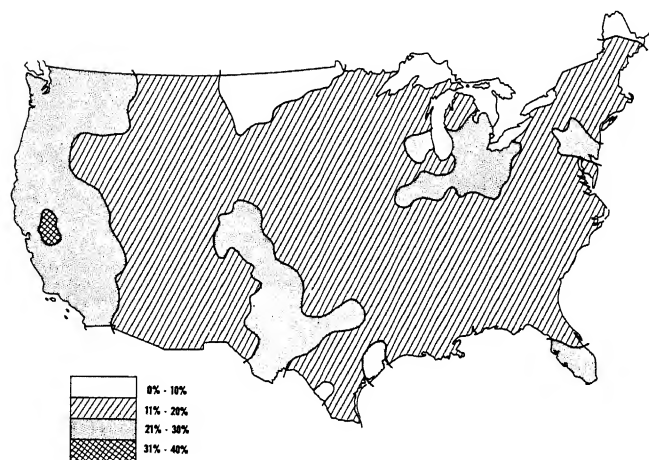


FIGURE 25. Market penetration by colour TV in 1967

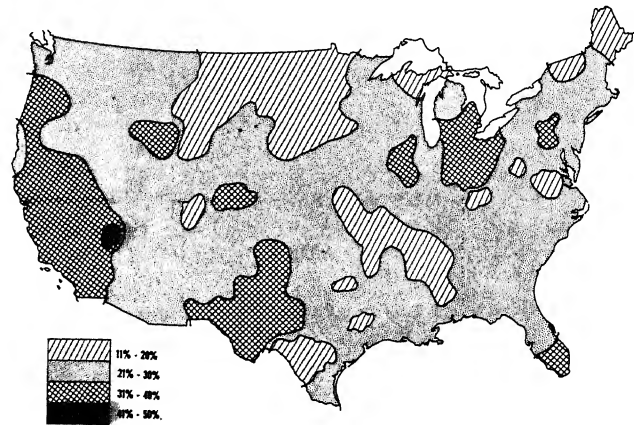


FIGURE 26. Market penetration by colour TV in 1968

homes—there will be more than 10 000 in only three or four years. And, following the vacation homes, will surely be the permanent residences and the industries.

I do not pretend for one minute that the name of the game is simply 'scatteration'. The essence of the change is that we are moving into an era of *telemobility*, and from mechanical into electronic environments. Our present geography and our geographical concepts are mechanical—distance-decay, gravity model, heartland-hinterland. To be sure, existing electronic technologies, if used in conventional ways, can so reduce the frictions of space and the delays of time that the traditional processes moulding today's geography will take on their limiting forms. But this again is the conservative, evolutionary view. The revolutionary aspect of electronic environments is not that they reduce the frictions in moving goods and people, but that they move the experience itself to the human nervous system. Traditionally, we have moved the body to the experience; increasingly we will move the experience to the body, and the body can therefore be located where it finds the non-electronic experiences most satisfying. Consider what is implied:

'Let us imagine a particular space and time *c.* 1986: a home in the suburbs of Phoenix. A man is sitting in the middle of a circular room and on the curved walls around him he can see the ocean-surf breaking over the rocks and foaming up the beach; a fish hawk trembling in the luminescent sky. Across from him sits another man, and the two of them are talking to each other. Once in awhile, the boom of the bursting surf and the cry of the hawk intrude upon their conversation.

'Let us now say that the room is underground and has no "real" view at all; that what is experienced on the curved walls is an image on a "flat wall" television screen, pre-recorded in Hawaii, and now being replayed electronically. Let us further say that the first man is "real", but that the second man is being broadcast by laser beam from a satellite and recreated, in color and full dimension (you could walk around his image and see the back of his head) by "holography", so that though he is "there" in Phoenix at the moment, he is "in reality" at the same moment sitting in his study at the University of Edinburgh.

'Where, in this situation, does "reality" begin and end? This will be a question that—by 1986—we will, individually, be asked to answer. There is nothing in the situation just described that does not appear to be perfectly feasible within perhaps the next ten years; certainly within the next twenty. We have *already* entered a new world of experience.'

What then of the current surge of growth of hotel space, meeting rooms and industrial plants around our major regional airports? What of corporation headquarters, of daily tides of commuters, of expressways and mass transit systems, if communication is substituted for movement? What if the geography of face-to-face contacts, of physical movement, of skyscrapers is replaced by a thin film of electrons spread over the countryside? Where then, did Boswash, Chipitts and Sansan go? Is this the most probable future? Is it the most desirable?

It is, I suppose, a failing that we share with all our species that we become wedded to comfortable ideas and retain them long after they have ceased to be relevant. So it is with current geographies of the United States, and even with views of its future. To be sure, images of past geographies peep through, palimpsest-style, as they will in the future, for Kenneth Boulding properly emphasizes that growth creates form, but that form then limits growth. But bold new patterns and forces lead us to re-fashion the picture we paint of the present, and the story we tell of the future. The signs are there for the discerning observer to see. The problem is to fit them into a pattern, to determine the process, to work out the logical consequences, and to outline the alternatives for achieving a different future if we do not like what we foresee, for as Daniel Bell argues so cogently (*personal commun*) 'perhaps the most important social change of our time is the emergence of a process of direct and deliberate contrivance of change itself. Men now seek to anticipate change, measure the course of its direction, and even shape it for predetermined ends'. And therein, I suspect, lies the future of geography as a profession, of the objects that it studies, and of its conceptual foundations.

#### NOTES

1. The television viewing areas used are the Designated Market Areas of the A. C. Nielsen Company (1967-68), defined such that each D.M.A. contains:

(i) The central metropolitan area for which at least one station achieves the largest quarter-hour audience share 9.00 a.m.-12.00 midnight.

(ii) All additional counties that have their largest quarter-hour audience share with all the D.M.A. stations in total. Counties are allocated exclusively to a particular D.M.A., and the definition is revised each year. Smaller single-station markets within larger D.M.A.s are not distinguished and, in the county allocations, boosters, C.A.T.V., satellites and translators are taken into account.

2. Starting with Figure 2, these were defined by the Regional Economics Division of the Office of Business Economics, U.S. Department of Commerce, as follows:

(i) Economic centres were identified. Usually, these were S.M.S.A.s but, where several S.M.S.A.s are parts of larger integrated economic complexes, they were combined. In rural parts of the country, centres of 25 000-50 000 were used, provided they were wholesale trade centres.

(ii) Counties were allocated to centres first on the basis of commuting flows, and for the periphery on the basis of other ties—television viewing, wholesale trade, etc.

(iii) Very small economic areas were combined if they comprised larger television or wholesale markets, to satisfy a minimum size criterion of 200 000.

The intent was to define areas that 'approach closure with respect to residuary industries, and therefore self sufficiency in the tertiary sector, while specializing in export activities to other areas' (*U.S. Department of Commerce*, 1967).

3. I am indebted to Calvin L. Beale for letting me draw on his maps in preparing Figures 8-11.

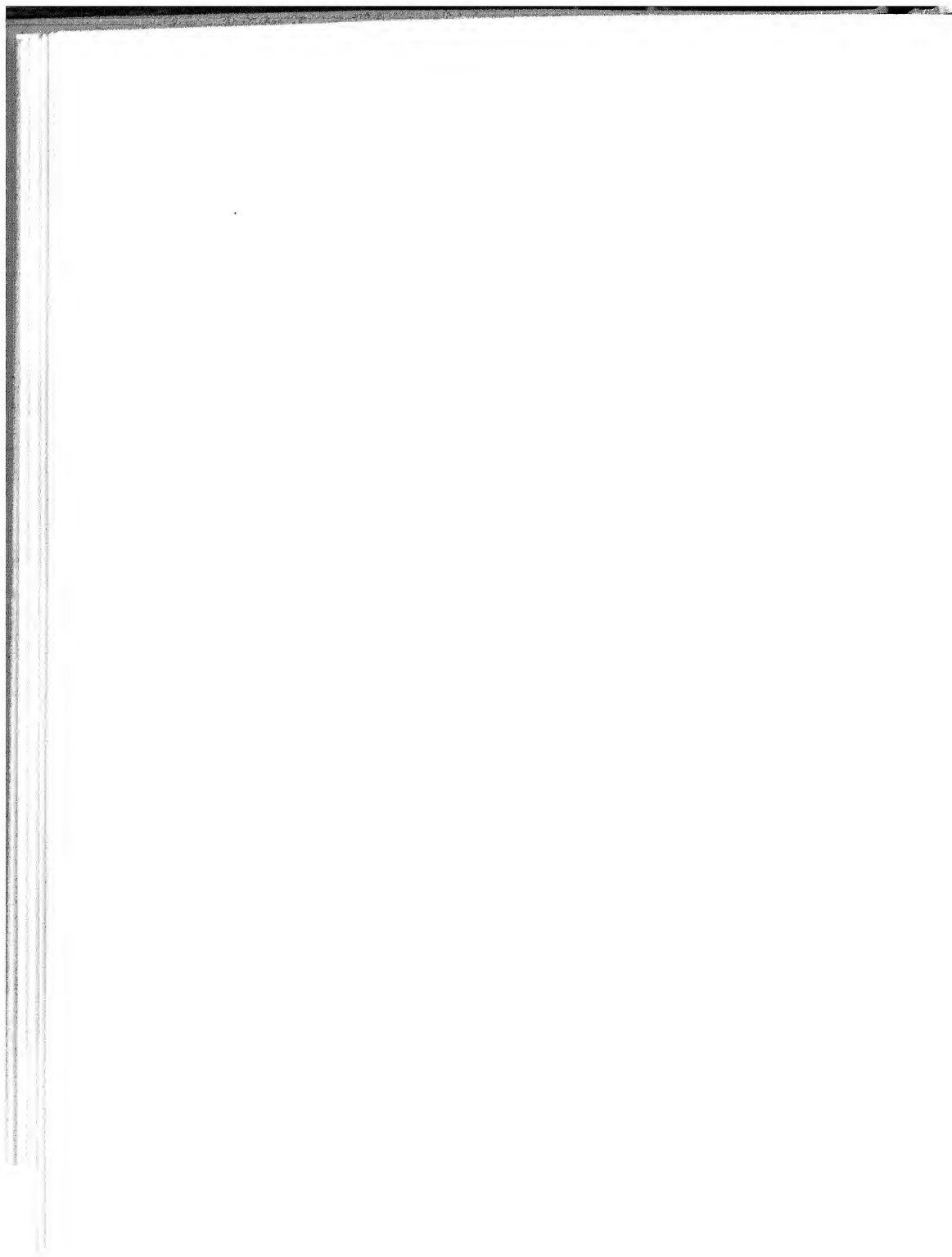
4. This quotation comes from a brochure of the Kaiser Company circulated several years ago called 'Telecom-bility—when Far is Near'.

## REFERENCES

- BERRY, B. J. L. (1970) 'Hierarchical diffusion: the basis of developmental filtering and spread in a system of growth centers', forthcoming in N. HANSEN (ed.), *Growth centers in regional development* (New York)
- BERRY, B. J. L. and F. HORTON (1970) *Geographic perspectives on urban systems* (Englewood Cliffs, N.J.)
- BORCHERT, J. R. (1967) 'American metropolitan evolution', *Geogr. Rev.* 57, 324
- BORTS, G. H. (1967) *Patterns of regional economic development in the United States* (Report to the National Advisory Commission on Rural Poverty, U.S. Department of Agriculture)
- DOXIADIS, C. A. (1969) 'Ekistics. An attempt for a scientific approach to the problems of human settlements', Paper presented to the Panel on Science and Technology of N.A.S.A.
- FRIEDMANN, J. (1966) *Regional development policy* (Cambridge, Mass.)
- FRIEDMANN, J. and J. MILLER (1965) 'The urban field', *J. Am. Inst. Plann.* 31, 312-19
- KAHN, H. and A. J. WIENER (1967) 'The next thirty-three years: a framework for speculation', *Daedalus* 96, 705-32
- KUHN, T. S. (1962) *The structure of scientific revolutions* (Chicago)
- NEILS, E. (1969) 'The urbanization of the American Indian and the Federal program of relocation assistance', Unpubl. M.A. thesis, Univ. of Chicago
- NIELSEN, A. C. Co. (1968) *Designated market areas* (New York, A. C. Nielsen Co.)
- PERLOFF, H. S. (1960) *Regions, resources and economic growth* (Baltimore)
- PICKARD, J. P. (1967) *Dimensions of metropolitanism* (Washington, D.C.; *Urban Land Inst. Res. Monogr.* 14)
- PRED, A. (1966) *The spatial dynamics of U.S. urban-industrial growth, 1800-1914* (Cambridge, Mass.)
- SERVAN-SCHREIBER, J. J. (1969) *The American challenge* (Harmondsworth, Middx.)
- THOMPSON, W. (1968) 'Internal and external factors in the development of urban economies' in H. S. PERLOFF and L. WINGO (eds) *Issues in urban economics* (Baltimore)
- ULLMAN, E. (1957) *American commodity flows* (Seattle)
- U.S. Department of Commerce (1967) 'OBE economic areas of the United States' (Washington, D.C.)
- VAUGHAN, R. (1843) *The age of great cities* (London)
- WELLS, H. G. (1902) *Anticipations of the reaction of mechanical and scientific progress upon human life and thought* (New York)

---

## **Spatial Systems in Economic Development**



## 6 Interregional and International Transmission of Economic Growth

Albert O. Hirschman

### *"Growing Points" and Lagging Regions*

TO COMPLETE our survey of inducement mechanisms, we shall examine in this chapter how growth can be communicated from one region or one country to another. In this inquiry we may take it for granted that economic progress does not appear everywhere at the same time and that once it has appeared powerful forces make for a spatial concentration of economic growth around the initial starting points. Why substantial gains may be reaped from overcoming the "friction of space"<sup>1</sup> through agglomeration has been analyzed in detail by the economic theory of location. In addition to the locational advantages offered by *existing* settlements others come from nearness to a *growing* center where an "industrial atmosphere" has come into being with its special receptivity to innovations and enterprise. It was largely the observation of the latter connections that suggested to Marshall the concept of external economies.<sup>2</sup>

Whatever the reason, there can be little doubt that an economy, to lift itself to higher income levels, must and will first develop within itself one or several regional centers of economic strength. This need for the emergence of "growing points" or "growth poles"<sup>3</sup> in the course

1. This term was used by Robert M. Haig in "Toward an Understanding of the Metropolis," *Quarterly Journal of Economics*, 40 (1926), 184-5.

2. A good survey of Marshall's views and of other contributions to this subject is in Eric A. Lampard, "The History of Cities in the Economically Advanced Areas," *Economic Development and Cultural Change*, 3 (Jan. 1955), 81-137, particularly 92-101.

3. "Pôle de croissance" is the term used for both regional and sectoral growth leadership in the expanding and instructive French literature on the subject. See, e.g., Perroux, "Note sur la notion de 'pôle de croissance'"; *Matériaux pour une analyse de la croissance économique*, Cahiers de l'Institut de Science Eco-



of the development process means that international and interregional inequality of growth is an inevitable concomitant and condition of growth itself.

Thus, in the geographical sense, growth is necessarily unbalanced. However, while the regional setting reveals unbalanced growth at its most obvious, it perhaps does not show it at its best. In analyzing the process of unbalanced growth, we could always show that an advance at one point sets up pressures, tensions, and compulsions toward growth at subsequent points. But if all of these points fall within the same privileged growth space, the forces that make for transmission of growth from one country, one region, or one group of persons to another will be singularly weak.

The ability and tendency of growth to round itself out for a long time within some subgroup, region, or country while backwardness retains its hold elsewhere has often been noted. If the tendency manifests itself along clearly marked geographic lines, the result is the division of the world into developed and underdeveloped countries and the split of a country into progressive and backward regions. On the other hand, progress and tradition may dwell in close spatial proximity by simply fastening on different human groups and economic activities that exist side by side; this state of affairs, often encountered in developing countries, has been aptly termed "dualism" and has already been examined in our analysis of the industrialization process (Chapter 7).

With respect to different social or income groups a similar phenomenon may be noted: once one group has shown its readiness to acquire new wants and its ability to afford the products satisfying them, it will be catered to by a multitude of firms all tailoring their output to the type of per capita buying power and to the size of the market that have been revealed. It takes innovators like Ford and Giannini to strike out beyond this charmed circle, just as it seems to take a special kind of boldness to establish a new basic industry or to perceive the development potentials of the more backward regions of a developing country.

Thus investors spend a long time mopping up all the opportunities around some "growth pole" and neglect those that may have arisen or

nomique Appliquée, Série D, No. 8, 1955; J. R. Boudeville, "Contribution à l'étude des pôles de croissance brésiliens," Cahiers, Série F, No. 10, 1957.

could be made to arise elsewhere. What appears to happen is that *the external economies due to the poles, though real, are consistently overestimated by the economic operators.*

The reason for this tendency—perhaps implicit in the phrase “nothing succeeds like success”—must be sought in the realm of social psychology. The progressive sectors and regions of an underdeveloped economy are easily overimpressed with their own rate of development. At the same time, they set themselves apart from the less progressive operators by creating a picture of the latter as lazy, bungling, intriguing, and generally hopeless. There seems to be a cliquishness about progress when it first appears that recalls the same phenomenon among adolescents: the girls who menstruate and the boys who shave have an acute sense of their superiority over those who cannot yet claim such achievements. The tendency to magnify the distance that separates one group or region from another shows up in the derogatory use of the term “indio” in some Latin American countries to designate whoever is economically or socially one’s inferior. Similarly, the average Italian, in whose country economic progress has long been closely associated with latitude, is always ready to declare that Africa begins just south of his own province.

Thus the successful groups and regions will widely and extravagantly proclaim their superiority over the rest of their country and their countrymen. It is interesting to note that to some extent these claims are self-enforcing. Even though the initial success of these groups may often be due to sheer luck or to environmental factors such as resource endowment, matters will not be left there. Those who have been caught by progress will always maintain that they were the ones who did the catching; they will easily convince themselves, and attempt to convince others, that their accomplishments are primarily owed to their superior moral qualities and conduct. It is precisely this self-righteousness that will tend to produce its own evidence: once these groups have spread the word that their success was due to hard work and virtuous living, they must willy-nilly live up to their own story, or at the least will make their children do so.<sup>4</sup> In other words, there is reason to think that the “protestant ethic,” instead of being

4. Observation would seem to confirm that the behavior of second generation businessmen is far more compulsively “businesslike” than that of the pioneer generation.

the prime mover, is often implanted *ex post* as though to sanctify and consolidate whatever accumulation of economic power and wealth has been achieved. To the extent that this happens, a climate particularly favorable to further growth will actually come into existence in the sectors or regions that have pulled ahead, and this will confirm the economic operators in their preference for these regions and make it somewhat less irrational.

The less developed groups and regions also make unwittingly a contribution to the process which we can only sketch here. Faced with the sudden improvement in the fortunes of some of their own compatriots, they will frequently retort to the claims of superiority of these *nouveaux riches* by accusing them of crass materialism, sharp practices, and disregard for the country's traditional cultural and spiritual values. While such charges are directed with particular relish at minorities, whose importance in the process of development is well recognized, purely indigenous entrepreneurial groups are by no means exempt from them. In this way these groups are, as it were, converted into minorities in their own country,<sup>5</sup> often estranged from the rest of their compatriots, and ostracized by the traditional elites. Such a development is particularly likely when the first stages of commercial and industrial progress are localized in a center other than the capital city. In this case, the rift between this center and the capital may well widen cumulatively over a long period of time. The very fact that the leading families of such cities as Barcelona, São Paulo, Medellín, and Guayaquil lived far away from, and often in conflict with, the centers of politics, public administration, and education made for a dogged concentration of succeeding generations on business pursuits rather than for absorption of the most talented by other careers that carry more prestige in a traditional society. This situation may again lead

5. A good example is supplied by the inhabitants of Antioquia, a province of Colombia. The Antioqueños have been outstandingly enterprising in bringing virgin lands under coffee cultivation and in establishing industries, mostly in their capital of Medellín. Their racial, religious, and cultural characteristics do not differentiate the Antioqueños from the other Colombians; but having taken such a prominent part in the country's development, they are *now* considered practically as a separate group; and even though it is unsupported by any evidence (see J. J. Parsons, *Antioqueño Colonization in Western Colombia* [Berkeley, 1949], p. 62), the conviction is widespread that they are really of Jewish, or at least of Basque, origin!

to a clustering of investment around the initial growing point, which is healthy for the consolidation of economic growth at its beginning but may represent irrational prejudice and clannishness at a later stage.

*Trickling-Down and Polarization Effects* <sup>5a</sup>

No matter how strong and exaggerated the space preference of the economic operators, once growth takes a firm hold in one part of the national territory, it obviously sets in motion certain forces that act on the remaining parts. In examining these direct interactions, we shall call "North" the region which has been experiencing growth and "South" the one that has remained behind. This terminology is suggested by the fact that a large number of lagging areas, at least in the Northern Hemisphere, appear to be located in the southern parts of the countries to which they belong. The term "South" as used here does not include *undeveloped*—i.e., largely unsettled—areas.

The growth of the North will have a number of direct economic repercussions on the South, some favorable, others adverse. The favor-

5a. Footnote added in proof: The argument of the following sections was outlined originally in my article "Investment Policies and 'Dualism' in Underdeveloped Countries," *American Economic Review*, 47 (Sept. 1957), 550-70. I now find that Gunnar Myrdal has addressed himself to similar problems in *Economic Theory and Under-Developed Regions* (London, 1957), particularly in chapters 3 to 5, and has had recourse to the same conceptual tools that are employed here: his "backwash" and "spread" effects correspond exactly to my "polarization" and "trickling down" effects. Nevertheless, there are considerable differences in emphasis and conclusions. Myrdal's analysis strikes me as excessively dismal. In the first place, he fails to recognize that the emergence of growing points and therefore of differences in development between regions and between nations is inevitable and is a condition of further growth anywhere. Secondly, his preoccupation with the mechanism of cumulative causation hides from him the emergence of the strong forces making for a turning point once the movement toward North-South polarization within a country has proceeded for some time. Finally, the picture he paints of international transmission of growth is also too bleak in my opinion as he overlooks that the polarization (backwash) effects are much weaker between nations than between regions within the same country. However, I fully agree with Myrdal on the importance of political forces in effecting a North-South rapprochement within a country and on the need for the emergence of such forces on the international level to help narrow the gap between the developed and the underdeveloped countries. I have anticipated here the discussion contained in the remainder of this chapter.

able effects consist of the *trickling down* of Northern progress: by far the most important of these effects is the increase of Northern purchases and investments in the South, an increase that is sure to take place if the economies of the two regions are at all complementary. In addition, the North may absorb some of the disguised unemployed of the South and thereby raise the marginal productivity of labor and per capita consumption levels in the South.

On the other hand, several unfavorable or *polarization* effects are also likely to be at work. Comparatively inefficient, yet income-creating, Southern activities in manufacturing and exports may become depressed as a result of Northern competition. To the extent that the North industrializes along lines in which there is no Southern production, the South is also likely to make a bad bargain since it will now have to buy Northern manufactures, produced behind newly erected tariff walls, instead of similar goods previously imported from abroad at lower prices.

A most serious, and frequently observed, polarization effect consists in the kind of internal migration that may follow upon the economic advances of the North. Instead of absorbing the disguised unemployed, Northern progress may denude the South of its key technicians and managers as well as of the more enterprising young men. This type of migration may actually be undesirable not only from the point of view of the South but also from that of the country as a whole, for the loss to the South due to the departure of these men may be higher than the gain to the North. This possibility is inherent in the contact between the expanding North and the stagnant South: in the North new jobs must be manned and, at least in the skilled grades, the wage and salary scale will reflect relative scarcities and productivities; whereas in the South skilled work and better-than-average performance will often be poorly remunerated either because they are simply not recognized or because they are not valued very highly<sup>6</sup> or because they carry nonmonetary rewards. Thus actual pay differentials between North and South are likely to overstate considerably the real productivity differentials in the most productive and skilled grades. In addition, of course, mobility is highest in these same lines so that it becomes almost a certainty that the South will lose to the North first

6. Even societies that actively discourage better-than-average performance (see p. 12) are unable to abolish it altogether simply because of innate differences.

and foremost its more highly qualified people. And, along with skill and enterprise, what little capital the South generates is also likely to migrate northward.

In spite of this bleak picture, we would still feel confident that in the end the trickling-down effects would gain the upper hand over the polarization effects if the North had to rely to an important degree on Southern products for its own expansion. For instance, if the North specializes in manufactures and the South in primary production, the expanding demand of the North ought to stimulate Southern growth. But things may go less smoothly. It is likely, in particular, that the short-run supply elasticity is low in the South so that the terms of trade will move against the North.<sup>7</sup>

In this case, three possibilities arise. In the best of worlds, the rise in Southern prices would fairly soon prove effective in raising production. Another possible, though far less satisfactory, outcome would consist in the slowing down of Northern progress resulting from rising labor and material costs. But such a development is unlikely as long as the North is not entirely dependent on the South. The third possibility is therefore for the North to alter its method of procuring needed primary products. Faced with the upward trend in Southern prices and exasperated by the unreliability of Southern production, Northern operators may draw on imports from foreign areas or may replace Southern products by developing their own primary production. In this way, *checks to the trickling-down effects* may well come into play, and as a result the South could be left in a far worse backwater than before. For once the North possesses within itself a large and productive agricultural area or is able to supply its needs in primary products from abroad and through domestic synthetic production, the South will be largely cut off from beneficial contact with Northern development, while remaining exposed to the adverse polarization effects. Under these conditions—which are or were fairly typical of such backward regions as Brazil's Nordeste, Colombia's Oriente, and Italy's Mezzogiorno—the stage would be set for a prolonged split of the country into a progressive and a depressed area.

7. This situation has been fully analyzed by H. G. Johnson for the case in which a developing industrial country trades with a stagnant agricultural country; see his "Economic Expansion and International Trade," *Manchester School of Economic and Social Studies*, 23 (May 1955), 96–101.

Eventually, economic pressures to remedy such a situation are likely to assert themselves again. Industry will become congested in Northern cities and its expansion will be hampered by the insufficient size of the home market resulting from the depressed income levels in the South. Also, economic policy makers will be impelled to take a close look at Southern development potentials whenever balance-of-payments or other supply difficulties make it clear that the country is harming itself by its failure to utilize fully its Southern resources.

In other words, if the market forces that express themselves through the trickling-down and polarization effects result in a temporary victory of the latter, deliberate economic policy will come into play to correct the situation. Actually, of course, economic policy will be an important influence throughout the process. The nature of this influence will be analyzed presently.

#### *The Regional Distribution of Public Investment*

The most obvious manner in which economic policy affects the rates of growth of different parts of a country is through the regional allocation of public investments. Three principal patterns of allocation can be distinguished: dispersal, concentration on growing areas, and attempts to promote the development of backward areas.

In contrast to widespread impressions, the most pervasive tendency of governments of underdeveloped countries in making their investment decisions is not so much the obsession with one showpiece as the dispersal of funds among a large number of small projects scattered widely over the national territory.

While this pattern is *dominant* only in countries where dynamic economic growth has not yet taken hold, it can be said to exert a steady pull in practically all underdeveloped countries. The most obvious reason is that public investment decisions are easily the most political ones among the economic policy decisions taken by governments. Whether to build a road here rather than there, whether to construct a power plant that is to supply towns A, B, and C, rather than D, E, and F—these are questions that have decisive local political impact.

Thus, as all governments regardless of their democratic character desire and need support from all sections of the country, the tempta-



tion is strong to scatter the investment effort far and wide. Disconnected roads are built at many points; small Diesel power plants and aqueducts are installed in many towns; even low-cost housing programs which should obviously concentrate on relieving critical shortages and on slum clearance in the big cities are often similarly dispersed.

More fundamentally, the tendency toward wide dispersal of investment funds may be due to what was called in Chapter I the group-focused image of change, i.e., to the fact that economic progress is conceived as a force which ought to affect equally all members and sections of the community. Wherever this idea prevails, governments are unprepared and unwilling to make the choices about priorities and sequences that are the essence of development programs. When the feeling is widespread that something is wrong with even temporarily preferred treatment for some regions, the government may find it politically dangerous not to take this factor into account.

Finally, the dispersal pattern can be explained by certain shortages usually affecting underdeveloped countries. The elaboration of the many small projects into which public investment is typically split up when this pattern is dominant requires comparatively little engineering and planning talent, whereas the larger projects in electric power, transportation, or basic industry require far more such talent than is usually available to the government. This is why entirely too much has been made of the argument that development is held back not by the scarcity of funds, but by a scarcity of "bankable," i.e., well-conceived and engineered, projects. The question which should come first, the project or the funds, is really of the chicken-egg variety. Obviously funds can be spent only on clearly defined projects. But without definite expectations that funds—from domestic or foreign sources—will be forthcoming, the considerable cost of engineering and economic studies and the administrative effort required to gather the necessary staff and to obtain the assistance of foreign consultants will most likely not be undertaken. The promise of foreign funds—provided the studies prove the project feasible and worth while—is particularly important if this effort is to be made, as a large project usually results in one region's obtaining for the time being a substantial advantage over all others. This is an investment decision which a national government may find it difficult and imprudent to make unless it has the feeling—and



the excuse vis-à-vis the other regions—that international development capital is not to be had at all on other terms.

Moreover, the study and preparation of a large-scale project implies in itself—especially in countries where there is the rhetorical tradition of confusing the word with the deed, and the announcement of plans with their realization—a commitment to the region which is going to be principally benefited. Governments are therefore reluctant to start such studies unless they feel reasonably sure that they will be able to “deliver.” Unless they have assurances in this regard, they would be politically much better off to let sleeping projects lie.

The International Bank for Reconstruction and Development has often defended itself against charges of insufficient lending by the argument that there were not enough “bankable” projects available.<sup>8</sup> But in fact the Bank has frequently acted in accordance with the point of view just outlined—i.e., it has helped in the preparation of such projects by virtually committing itself in advance to the financing of their foreign exchange costs, including even the cost of the preliminary engineering surveys.

In this way the availability of international development capital may make for a shift from dispersal of public investment toward concentration on a few key projects. The “demonstration effect” of similar projects undertaken in other countries also works in this direction. But the most important force opposing the tendency toward excessive dispersal of public investment is the growth pattern characteristic of rapidly developing countries. Development often begins with the sudden, vigorous, and nearly spontaneous growth of one or a few regions or urban centers, resulting in serious shortages of electric power and water supply, as well as in housing and transportation bottlenecks. Thus, urgent demands for several types of capital-intensive public investment appear and must be given the highest priority whether or not

8. Statements to this effect can be found in several of the Bank's annual reports; e.g.: “Perhaps the most striking single lesson which the Bank has learned in the course of its operations is how limited is the capacity of the underdeveloped countries to absorb capital quickly for really productive purposes. . . . The Bank's experience to date indicates that the Bank now has or can readily acquire sufficient resources to help finance all the sound productive projects in its member countries that will be ready for financing in the next few years, that can appropriately be financed through repayable foreign loans and that cannot attract private capital.” *Fourth Annual Report* (Washington, 1948–49), pp. 8, 13.

they correspond to the government's sense of distributive justice and to its pattern of regional political preference. The public investment in overhead capital in turn makes possible further growth of industry and trade in the favored areas and this growth requires further large allocations of public investment to them.

Determined as it is by the volume of private investment and the general rise in income in the developing areas, public investment clearly plays here an "induced" role, and investment choices are often remarkably and unexpectedly obvious. It is not always easy, however, to have these obvious choices adopted, partly because of the continuing desire of governments to revert to the policy of scatter, and partly because a new pressure soon makes itself felt—namely, to accelerate development in the areas that have fallen behind.

A situation in which the bulk of public investment is continuously being sucked into the comparatively developed portions of the national territory cannot in the long run be considered satisfactory by governments because of compelling considerations of equity and national cohesion. In fact, the attempt to change drastically the distribution of public investment in favor of the country's poorer sections often comes at a point that seems premature to the foreign observer or adviser for the simple reason that the more rapidly advancing sections do not strike *him* as so outstandingly prosperous. It is, however, quite understandable that the attempt should be made long before these sections have come anywhere near fully developing their potential. Moreover, the poorer sections of the country, where careers in industry and trade are not promising, often produce, for this very reason, a majority of the country's successful politicians and thereby acquire influential spokesmen in the councils of government.

It is possible that the transition from the second pattern—concentration of public investment on spontaneously growing areas—to the third—attempt to ignite development in the heretofore stagnant areas through "autonomous" public investment—is facilitated by certain peculiar properties of public investment. Usually the second phase results not in a mere shift from scatter to concentration of a given investment total, but in a considerable enlargement of the total amount of funds required for public investment. These funds are secured through the introduction of new and higher taxes or through other *permanent* revenue-raising devices.

On the other hand, it is probably reasonable to assume that the need for the investment of public funds in the country's spontaneously growing areas is particularly great in the initial stages of development, as basic utilities are created and rapidly expanded. After development has proceeded for some time, the need for public investment in relation to private investment tends to decline and in any event an increased portion of public investment can be financed out of earnings of previous investments. This kind of change in the composition of investment is implicit in the term "social *overhead* capital."

As the taxation and other measures which have financed the original spurt in public investment continue to yield revenue, some funds may thus become, if not unemployed, at least less compellingly employed than previously. This is likely to be immediately sensed by the officials responsible for apportioning public investment and provides an excellent opportunity to those among them who want to change its geographic composition in favor of the less developed sections.

Thus, while public investment policy may accentuate at one stage the North-South split, it can be counted upon to stage at least an attempt to heal the split should it turn out to be prolonged. For this reason governmental intervention is particularly prevalent in the development of the backward areas within underdeveloped countries. In fact, the government will, to the best of its ability, attempt to counteract in part the polarization effects that result from the operation of market forces: to counterbalance the northward emigration of capital and talent, an even larger flow in the opposite direction will be organized; to offset the locational advantages of the North, governments may offer special tax advantages or create similar external economies in the South through public investments.

Naturally, the channeling of large-scale expenditures toward the underprivileged areas of a country contains the danger of misguided investment to a much higher degree than where spontaneous growth has already staked out fairly well the areas in which public investments are urgently required. The most obvious and least "risky" course<sup>9</sup> is to endow the South with just as good a system of transportation, electric power stations, and other social overhead capital facilities as are available in the North. But we have already explained that this may not be the most efficient method of inducing growth in

9. See pp. 165-6.

the South because of the weakness of its entrepreneurship and the purely "permissive" character of the inducement mechanisms set in motion by these investments. Although some investment in public utilities may be indispensable, the essential task is to endow the South with some ongoing and actively inducing economic activity of its own, in industry, agriculture, or services. For this reason, the building of a steel mill in Colombia's Oriente and the founding of the new Brazilian capital in the long neglected "interior" will probably turn out to be effective governmental development moves in spite of initial mistakes, difficulties, and setbacks.

### *Interregional and International Transmission Compared*

Our discussion has made it clear that the interregional transmission of growth cannot be expected to proceed smoothly. Obstructionist forces have been seen to be at work alongside those that make for integration, in the economic and political fields. It is tempting then to apply an *a fortiori* argument to the *international* transmission of growth: if interregional transmission is beset with obstacles, is it not natural to conclude that international transmission will be even more difficult?

While the disparity in the development levels of different countries would seem to support such a conclusion, it is not at all justified by the arguments we have used in demonstrating the difficulties of interregional transmission of growth. Some of these arguments rather point to the somewhat unsettling thought that the various "Souths" might be better off if they were sovereign political units, i.e., that in some respects growth may be more easily transmitted from one nation to another than from one region to another within the same country. We will first explore these "economic arguments for separatism" and then show in what respects transmission mechanisms are after all more effective between regions than between sovereign countries.

*The case for separatism.* In general it may be expected that because of the closer contact and more intensive interaction that exist among regions of the same country than among sovereign nations, both trickling-down and polarization effects will be found to be stronger in interregional than in international economic relations.

The case for separatism will therefore consist largely in showing that the polarization effects will be far less damaging to a country than to a region. This is certainly the case for the mobility of factors of production. We have seen that within a country this mobility can be highly prejudicial for the South, and conceivably even uneconomic from the point of view of the country as a whole. If the South were an independent country, mobility would certainly be far lower and the Southern development potential would be less impaired.

Another polarization effect consisted in the debilitating influence of Northern competition on Southern economic activities satisfying domestic or export demands. Again, this effect would be virtually absent between independent countries. With respect to the latter, countries compete in international markets on the basis of comparative advantage, regions within a country on the basis of absolute advantage. Suppose that North and South, considered independently, both have a comparative advantage in cane sugar, but that production is more efficient in the North. Then, if each were an independent country, they would both specialize in sugar, with real factor returns being lower in the South. But if North and South are united in one country, sugar production would be expanded in the North and may be abandoned in the South even though the maintenance and expansion of sugar exports could represent the valuable beginning of a "growth pole" for the South.

The same reasoning holds for industrialization. It has frequently been pointed out that, if there is any substance to the various arguments for protection, they must apply just as much to a region as to a country; but the region cannot ordinarily protect its industries except through exemption from minor local taxes. Also, within a country (or in relations between a country and its colonies) Northern industrialists may often effectively prevent or delay the development of industry in the South; in relations between sovereign countries, attempts in this direction have sometimes been made, but obviously have far smaller chances of success.

Finally—and related to the previous points—the absence of economic sovereignty with respect to such matters as currency issue and exchange rate determination may be a considerable handicap for the development of a region.

The preceding argument is reminiscent of Viner's celebrated thesis

that a customs union between two countries may lead to a less, rather than more, efficient allocation of resources.<sup>10</sup> To prove this proposition, Viner invoked only the "trade-diverting effects," i.e., the possibility that the partners of a customs union will now buy from each other what they could previously obtain more cheaply—and what can be more efficiently produced—in third markets. This argument is also applicable to our problem, but the polarization effects relating to factor mobility and North-South competition in exports and industry are perhaps more important in a developmental situation.

*The case for surrender of sovereignty.* We must now come to the other half of our story. As the polarization effects will be stronger when there are no frontiers to cross, so will the trickling-down effects. The advance of the North is bound to lead to purchases and investments in the South. All complementarities that exist within a country will be readily exploited. Regional specialization patterns will emerge and are not likely to be tampered with even when they are based more on historical accident than on comparative resource endowments. Not so between sovereign countries. Here potential complementarities are likely to be taken advantage of in a far more selective and spotty fashion, not only because of the "friction of space" but because of the many other frictions that are encountered as soon as frontiers are crossed. Protectionist movements and reactions to balance-of-payments difficulties will set up strong obstacles to the emergence of a finely articulated division of labor and will always threaten it if it should arise.

The trickling-down effects will still be powerfully effective in promoting development of countries with resources that are highly prized by the industrial countries. But if a country has nothing particularly essential or attractive to offer, it may remain excluded for a long time from any important participation in world trade when, as a region integrated into a larger country, it might have contributed quite nicely to interregional trade.

In our treatment of interregional transmission of growth we saw the principal danger of an emergence of a North-South problem in

10. Viner, *The Customs Union Issue* (New York, 1950), ch. 4. For a systematic discussion of the issues raised by Viner, see J. E. Meade, *The Theory of Customs Unions*, Amsterdam, 1955.

the low supply elasticity characteristic of the South and in the consequent loss of interdependence as the North extricated itself from dependence on Southern products in a variety of ways. In international relations these checks to the trickling-down effects are stronger, just as the trickling-down effects themselves are weaker, than in inter-regional relations. Within a country, the loss by the South of its markets in the North will be resisted: not entirely unselfish efforts will be made by Northern interests to help the South overcome its supply difficulties which, if unresolved, will make it necessary for the North to look elsewhere. And, as has already been pointed out, even if a temporary lapse in North-South trade occurs, such efforts are likely to be resumed whenever balance-of-payments or other supply difficulties press upon the country.

In relations between advanced and underdeveloped nations, one of the major forces making for the growth of the latter is the need of the advanced nations for certain, usually primary, products of the underdeveloped areas. But if the foreign producers for one reason or another are unable to fulfill the rapidly growing requirements of the industrial centers, they cannot expect to be treated with the same patience and periodic consideration that they would receive if they were part of the industrial countries themselves. Domestic or other foreign sources of supply will be tapped or synthetic production will be undertaken largely on the basis of economic calculations, whereas political and social considerations will importantly affect similar decisions in North-South relations and are likely to help the South retain its role as a supplier of the North.

In this fashion we are brought back to the political forces making for the transmission of growth. These forces help definitively to redress the balance of our argument away from separatism. Within a country, as we have seen, there will come a point when a determined effort will be made to pull the underdeveloped regions within that country out of their stagnation. The ultimate reason for the confidence one may have in the emergence of such an effort is the solidarity that binds different parts of a nation together and the ability of each part to make itself heard and to bring pressures to bear on the central government. In spite of much progress in recent years, international solidarity of this kind is unfortunately still in its infancy.

We conclude that, on balance, the forces making for interregional

transmission of growth are likely to be more powerful than those making for international transmission.

### *Optimal Institutional Arrangements*

The reader may wonder why we examined in so much detail whether it is better for an underdeveloped area to be a region or a nation. Few areas can choose. Nevertheless, the realization that growth is transmitted more easily between nations than between regions from the point of view of some of the mechanisms we have analyzed, while the opposite holds for others, makes it tempting to think about the possibility of optimal institutional arrangements. If only we could in some respects *treat a region as though it were a country* and in some others *treat a country as though it were a region*, we would indeed get the best of both worlds and be able to create situations particularly favorable to development.<sup>11</sup>

Let us look first at the regions. Their advantage consisted largely in their greater exposure to the trickling-down effects and in their ability to call for help from the larger unit to which they belong. Their disadvantage seemed to lie principally in their exposure to polarization effects, in their inability to develop production for exports along lines of *comparative* advantage, and in the absence of certain potentially development-promoting policy instruments that usually come with sovereignty. A nation attempting to develop its own backward regions should therefore provide certain "equivalents of sovereignty" for these regions. The most important of such equivalents is a reaction against the feelings of despondency and self-denigration so often encountered in the South, and the mobilization of its energies through regional institutions and programs. The need for this approach has been felt in several countries where regional development corporations and banks have been set up. Effective aid to the establishment of industries in the South may call also for national income tax deductions (equivalent to tariff protection) and some autonomy in bank credit policy. To permit production to proceed on the basis of com-

11. We assume that the areas we are talking about have a substantial untapped development potential. There are of course many regions and perhaps even some countries whose natural resources are so poor or depleted that their best hope lies in becoming empty spaces—or at least far emptier than they are now.



parative advantage, Southern exports could be—and have at times been—stimulated through preferential exchange rates. Under such conditions, it might be held that imports into the South should be subject to compensating surtaxes, but this complication can be avoided on the ground that the South could satisfy many of its needs more cheaply in world markets if it were not prevented from doing so by the protection of Northern industries.

It is in line with our analysis that a policy of closing the gap between the South and the North requires the use of instruments that would ordinarily be thought to be disruptive of the very integration they are designed to achieve. While it is the purpose of these instruments to cut down the strength of the polarization effects, great care must be taken, of course, not to interfere with the efficacy of the trickling-down effects. Thus, the economic policies just outlined are designed to insulate the South sufficiently so that it may undertake certain industrial and export activities in competition with the North; but, at the same time, the complementary relationships that make the South a supplier of the North must be preserved and intensified.

For *international* transmission of growth, the optimal institutional arrangements would be of the opposite kind. The task here is to keep the polarization effects as weak as they normally are among independent nations, but to increase the strength of the trickling-down effects. In other words, the underdeveloped countries ought to retain the developmental advantages of sovereignty: obstacles to the emigration of skills and capital and a measure of independence in tariff, monetary, and foreign exchange policy. At the same time, they must be more closely integrated into the world economy through arrangements that make for more rapid growth and greater stability in their export markets. In addition, their development could of course be greatly accelerated if the community of nations disposed of a political mechanism similar to the one that within a nation makes eventually for a redistribution of public investment funds in favor of the South.

The world is already groping for formulas that would combine in this way the advantages of sovereignty with those of integration.<sup>12</sup> For the time being, these efforts are largely the incidental results of

12. Attempts of sovereign countries to assess themselves for costs incurred in joint programs are reviewed in THOMAS C. SCHELLING, *International Cost-Sharing Arrangements*, Essays in International Finance No. 24, Princeton, 1955.

a struggle for power. Yet it is obvious that they would be intensified rather than abandoned if this struggle were to cease tomorrow. It seems a pity, therefore, that we in the United States insist so loudly that the bold and pioneering steps we are taking to help the underdeveloped countries are dictated by military necessity or are "straight-forward business transactions."<sup>13</sup> Must we thus pave with apologies the road to what can be one of mankind's highest achievements? But perhaps it is inevitable that progress along this road should be reluctant. For, as Bergson has said, "the moral, original and fundamental structure of man is made for simple and closed societies . . . man outwits nature when from the solidarity of these societies he steps into human fraternity."<sup>14</sup>

13. Gunnar Myrdal makes some interesting observations on this point in *An International Economy* (New York, 1956), ch. 9.

14. *Les deux sources de la morale et de la religion* (Paris, 1934), pp. 53-4.

# 7 Regional Inequality and the Process of National Development: A Description of the Patterns \*

Jeffrey G. Williamson

## I. Introduction

Economists have long recognized the existence and stubborn persistence of regional dualism at all levels of national development and throughout the historical experience of almost all presently developed countries. Increasingly active theoretical discussions, empirical research, and especially political concern with this aspect of economic growth has given the phenomena of regional imbalance and inequity a popular new term—the "North-South problem." In spite of the recent attention which this problem has attracted, very little progress has been made in formulating and testing a general explanation for the occurrence of inequality in the spatial distribution of national income. One only needs to observe that Frenchmen, Italians, Brazilians, and Americans still tend to treat their North-South problems<sup>1</sup> as unique to their own national experience with economic growth. This may be explained by the fact that only a small amount of research effort has been devoted to comparative studies of

---

\* A number of agencies and individuals have aided us in gathering the data which made this study possible. To all of them we give our thanks. It would be fitting at this point to express our gratitude especially to Richard A. Easterlin, Simon Kuznets, Colin Clark, John A. Brittain, and Alfred Conrad for their help and stimulating suggestions. Thanks should also go to the Social Science Research Council for supplying financial assistance for this study; to the Vanderbilt University Computer Center for their generous help; and to the Wisconsin Social Systems Research Institute for secretarial assistance.

1. Throughout this paper we use such terms as "North-South problem" and "regional dualism" interchangeably with regional income differentials. They are not to be interpreted literally, since in comparing regions there is a whole spectrum or range of regional differentials—not just a dichotomy. Furthermore, it must be obvious to the reader that North is not equivalent to developed for all nations. These are purely literary simplifications.

regional inequality as related to the process of national development.<sup>2</sup> This empirical investigation into the nature of spatial inequality within national borders and over the development spectrum is an attempt to fill that void. Unfortunately, only a description of the aggregate patterns is presented here. It must be frankly admitted that, to a large extent, the more difficult task of disaggregation and identification of causation is left untouched.

## II. Expectations

There is an abundant accumulation of theoretical writings in which hypotheses about the nature of regional inequality during the development process are implied. Given that significant economic growth first appears in one region of a national state, it should occasion no surprise that the *absolute differential* between rich and poor regions (North and South) should persist or even increase. Even if both regions should grow at the same percentage rate after the fortuitous "random shock" in the North, the absolute regional differential will not only persist but increase. Regional income differentials are measured in this paper, however, in terms of relatives, not absolutes: the income per capita of each region is taken as a percentage of the average national income per capita. For example, the Brazilian Northeast in 1959 contained 25 percent of Brazil's population but only 10 percent of her income. The Southern states, on the other hand, contained 35 percent of the population but 50 percent of the income. In a less awkward fashion, the degree of inequality may be better summarized by indicating that most of the Northeastern states had per capita incomes of less than 50 percent of the Brazilian national average.

An inequality measure of this sort implies a comparison of regional growth rates and is much more informative for our purposes than one which considers absolute differentials.<sup>3</sup> Using this measure as the most appropriate index, what *a priori* notions might we have about the behavior of regional income differentials as national development proceeds? Does our "historic and current system of social and economic organization [perpetuate] interregional growth and income differentials once they come into existence?"<sup>4</sup> The answer may be as easy, or as difficult, as explaining why growth tends to be high and self-sustaining in nations which have already experienced it and so difficult to generate in currently underdeveloped Asian and African nations. The increasing divergence in international income per capita levels, at least prior to World War II, is well known, and a similar theoretical apparatus may be used to

2. There are, of course, significant exceptions. Besides the increasing empirical evidence relating to well-known North-South problems in Italy and Brazil, there has been an active interest in regional inequality in American historical development. See, for instance, Richard A. Easterlin, "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950," in *Trends in the American Economy in the Nineteenth Century* (Princeton: Princeton University Press, 1960), pp. 73-140 (hereafter called *Trends*); and Frank A. Hanna, *State Income Differentials, 1919-1954* (Durham: Duke University Press, 1959). Furthermore, the *Economic Survey of Europe in 1954* (Geneva, 1955), Ch. 6, pp. 136-71, devoted a good part of that issue to an examination of regional imbalance and inequality within the European nations. For the most recent examples of studies of this sort see José Raymon Lasuen, "Regional Income Inequalities and the Problems of Growth in Spain," *Regional Science Association Papers*, VIII (1962), 169-88; Minoru Tachi, "Regional Income Disparity and Internal Migration of Population in Japan," *Economic Development and Cultural Change*, XII, No. 2 (January 1964), 186-204; Werner Baer, "Regional Inequality and Economic Growth in Brazil," *Economic Development and Cultural Change*, XII, No. 3 (April 1964), 268-85.

3. The problem of choice of indices is clearly an important one; we discuss this point below at length.

4. R. B. Hughes, "Interregional Income Differences: Self-Perpetuation," *Southern Economic Journal*, XXII (July 1961), 41.

predict increasing divergence among geographic units *within* national borders and perpetuation of "poles de croissance."<sup>5</sup>

But presumably economic interdependence among regional units within nations should be much stronger than between countries themselves. Retaining the most restrictive classical assumptions, internal factor mobility should tend to eliminate interregional income per capita differentials, geographic dualism, or spatial polarization. Under conditions of free factor mobility, and abstracting from transportation costs, spatial inequality can persist only via lags in dynamic adjustment. That spatial inequality, depressed areas, and backward regions appear to persist may simply suggest to some that internal factor flows (tending to reduce interregional inequality) do not occur with sufficient speed and quantity to offset the dynamic indigenous conditions which cause relatively faster resource augmentation and technological change in the rich developing regions (tending to increase inequality).

In fact, one could reasonably appeal to the high degree of sectionalism, fragmentation, and general national disintegration in the youthful stage of national development to predict increasing regional inequality during those early decades. Given that young nations historically, as well as those currently, embarking on modern economic development have been typically devoid of national labor, capital, and trade markets approaching even rudimentary degrees of efficiency, this seems the only reasonable prediction. Regions within nations do not typically possess equal capacity for growth, and when development begins in some of these islands, regional barriers may be too great to communicate the growth stimulus to other less fortunate regions. As long as the barriers to trade and factor flows (as well as communication of technological change) persist, regional inequality will clearly increase.

The problem is hardly that simple, however. Myrdal's theorizing about backwash effects, Hirschman's concern with dualism and polarization, and Kuznets' more cautious "empirical" guesses suggest that even internal factor flows may not always be equilibrating in the classical fashion. On the contrary, in the initial stages of national development regional inequality is likely to increase all the more sharply due to a number of disequilibrating effects.<sup>6</sup>

#### A. Labor Migration

Interregional labor migration is likely to be extremely selective because of either the prohibitive money costs of migration at low levels of income or traditional inertia in the non-urbanized, non-industrialized poor Southern regions. The migrants may be characterized as the vigorous and entrepreneurial, the educated and skilled, and of productive age. (We are not describing the dominant characteristic of emigrants from the backward South, but suggesting that these characteristics will be more prevalent among the migrants than among the average population of the Southern regions.) Selective migration of this type obviously accentuates the tendency towards regional income divergence: labor participation rates, *ceteris paribus*, will tend to rise in the rich and fall in the poor regions; furthermore, precious human

5. This is a term used often in French literature to describe regional growth differentials. See F. Perroux, "Note sur la notions de 'pole de croissance,'" *Cahiers de L'Institut de Science Economique Appliquée*, Series D, No. 8 (1955), and Hirschman's use of the derivative "polarization" in his *The Strategy of Economic Development* (New Haven: Yale University Press, 1958), Ch. 10. It should be pointed out that the efforts of Perroux and his students have not added much to our knowledge of the process of interregional communication of growth. See Charles P. Kindleberger, *Economic Growth in France and Britain, 1851-1950* (Cambridge: Harvard University Press, 1964), pp. 259-60.

6. Disequilibrium here describes an internal factor flow which tends to increase regional inequality. We are not necessarily using it to describe factor movements which do not respond correctly to interregional factor price differentials. The two concepts may or may not coincide.

capital will tend to flow out of the South and into the North, making regional resource endowment per capita all the more lopsided and geographic imbalances all the more severe.

What has been said above about migration patterns in early development stages is hardly original to this study. It appears as one important theoretical buttress for the operation of both Myrdal's backwash effects<sup>7</sup> and Hirschman's polarization effects:

Instead of absorbing the disguised unemployed, Northern progress may denude the South of its key technicians and managers as well as of the more enterprising young men. . . Thus actual pay differentials between North and South are likely to overstate considerably the real productivity differentials in the most productive and skilled grades. In addition, of course, mobility is highest in these same lines so that it becomes almost a certainty that the South will lose to the North first and foremost its more highly qualified people.<sup>8</sup>

To illustrate the potential disequilibrating effects of interregional labor migration, Eckaus has characterized migration patterns during the early stages of Italian industrial development as being precisely of this type, although his contention is not defended empirically:

The nineteenth century was a time of extraordinary emigration, generally heavier in the South, where at some times and places it exceeded the natural increase in population. The concentration of immigrants in the younger, productive ages left the South with a working force of deteriorating quality.<sup>9</sup>

Dziewonski presents us with a contemporary example of the perverse effects of interregional migration at low levels of national development. He has noted that the Polish government has deliberately minimized interregional labor migration, since central planners are concerned with the past effects it has had in further contributing to regional inequality and dualism via effects upon relative participation rates and labor force quality.<sup>10</sup> This, too, appears to be consistent with the notion that in early stages of growth interregional labor flows may generate further spatial inequalities rather than reduce them. This would appear to follow from the fact that Poland has not reached a mature stage of growth.

#### B. Capital Migration

The interregional flow of private capital may tend to be perverse as well. External economies and general benefits derived from agglomeration of capital projects in the relatively rich Northern regions may cause capital to emigrate from the South to the North, tending to accelerate interregional inequality and to widen the North-South schism. High apparent risk premiums, lack of entrepreneurial ability, and immature capital markets may further depress investment activity and capital accumulation in the South. The latter, immature development of financial institutions, may prove to be not only important but also the most easily measurable of these factors in explaining perverse capital flows. Spain may serve as our example here:

7. Gunnar Myrdal, *Economic Theory and Underdeveloped Regions* (London, 1957), Ch. 3-5.

8. Hirschman, *op. cit.*, pp. 188-89. The Canadian province of Nova Scotia was for a long time noted for its major exports of "brains" and fish.

9. Richard S. Eckaus, "The North-South Differential in Italian Economic Development," *Journal of Economic History* (September 1961), 317.

10. Kazimierz Dziewonski, "Theoretical Problems in the Development of Economic Regions," *Regional Science Association Papers*, VIII (1962), 47.

Capital migrates mainly through the banking system. Spanish banks are highly of a mixed character, being commercial and industrial, and are highly oligopolistic. Seven banks handle more than seventy percent of the total credit. The result is that the deposits of the backward regions are transformed into credits for the industries in the north, particularly for those industries in which the banks participate. But capital migrates also via the capital market, for benefits are more certain and higher in the developed industrial sectors of the country. Most of the direct investment by entrepreneurs of Southern origin is also made in developed regions. Better infrastructural setting, superior transport and communications facilities, and larger markets all play a role.<sup>11</sup>

Nor is the evidence of perverse interregional flows of private capital isolated to the underdeveloped nations of Europe. The same pattern appears to exist in Pakistan, with heavy capital flows from East to West Pakistan, and in Indonesia, with similar flows from the outer to the central islands. And, of course, given our accumulated evidence that capital flows are heavily influenced by growth rates (demands for capital), capital "scarcity" in the South does not always imply high marginal productivity and high price.

#### C. Central Government Policy

The national or federal government's overt or unconscious intention to maximize national development may tend to increase still further the degree of regional inequality if active political expression in the South is lacking (as in the American post-Civil War Reconstruction period) or even in spite of such expression. In an overt fashion, the central government may allocate investment to the North, where "urgent demands for several types of capital-intensive public investments appear,"<sup>12</sup> which favors the fast-growing industrial regions and helps generate even more rapid growth there, only to create large social overhead capital requirements in the future. This is a common argument in the historical arsenal of Southern apologists in both Italy and the United States to explain relative lags in Southern industrial development.

In a less overt but equally important fashion, the central government may manipulate the external terms of trade in favor of the industrial North. A national tariff policy implemented with the intention of fostering and protecting industrial development, common to most underdeveloped nations past and present, clearly involves a geographic transfer to the rich Northern regions.<sup>13</sup> Southern senators and representatives in the United States were certainly aware of this and attached great importance to its effect during the three or four decades prior to the Civil War when their voice was important in American policy making.

#### D. Interregional Linkages

More generally, there may be a lack of interregional linkages in the early stages of national growth, so that the spread effects of technological change, social change, and income multipliers are minimized. Part of the national growth process is simply economic unification of regional markets. To the extent that such interregional linkages are slow in developing,

11. Lasuen, *op. cit.*, pp. 179-80.

12. Hirschman, *op. cit.*, p. 192.

13. Eckaus questions the significance of a protective tariff policy in contributing to 19th and early 20th century Italian North-South differentials. *Op. cit.*, pp. 313-14.

national development is all the more likely to be regionalized in the earliest stages of growth. Furthermore, if the North possesses a large and productive agricultural area, "the South will be largely cut off from beneficial contact with Northern development, while remaining exposed to the adverse polarization effects."<sup>14</sup> This factor should help explain the relatively severe problems of North-South dualism which have persisted, for example, in the histories of Brazil's Nordeste, Colombia's Oriente, Italy's Mezzogiorno, and the U.S. South.

\* \* \*

The working hypothesis of this study, however, is not that interregional divergence of income per capita levels will persist indefinitely into the mature stages of national growth. On the contrary, there are a number of reasons why we should expect the elements which tend to cause divergence to diminish over time, allowing the more classical equilibrating effects to make themselves felt.<sup>15</sup>

#### E. Labor Migration

Migration is likely to become less selective as economic development proceeds. There are a number of justifications for this expectation. Growth will have been occurring in the poor regions, although at a slower rate, and the prohibitive costs of migration may disappear, eliminating the bias against the unskilled and low-income groups in the Southern regions. Traditional rural inhibitions to interregional migration should have been significantly weakened by whatever economic progress has occurred in the South. Occupational wage differentials between the skilled and the unskilled are likely to diminish in the North relative to the South, further causing a change in the composition of internal migration. The South may not only retain its educated and skilled, while losing its unskilled, but may even attract the former type of migrant from the North.<sup>16</sup> At any rate, it certainly seems likely that the rate of internal labor mobility should increase as the integration of regional markets into a truly national economy proceeds. This has been the case historically for most developing nations: even after the Turnerian frontiers were filled in the United States in the 1890's, population mobility refused to decline and even increased in recent decades;<sup>17</sup> the same appears to have been the case in France, since the degree of population mobility has increased throughout the late nineteenth and early twentieth centuries.<sup>18</sup>

14. Hirschman, *op. cit.*, p. 189. See also John Friedman's discussion of regional dualism in "Regional Planning: A Problem in Spatial Integration," *Regional Science Association Papers*, V (1959), 167-79.
15. "Myrdal's analysis strikes me as excessively dismal. In the first place, he fails to recognize that the emergence of growing points and therefore of differences in development between regions... is inevitable and is a condition of further growth anywhere. Secondly, his preoccupation with the mechanism of cumulative causation hides from him the emergence of the strong forces making for a turning point once the movement toward North-South polarization within a country has proceeded for some time." Hirschman, *op. cit.*, p. 187.
16. What little evidence we do have suggests that this is certainly the case of the American South in the post-World War II period. Just how far back in American economic history this pattern can be traced is uncertain.
17. See Stanley Lebergott, *Manpower in Economic Growth: The United States Record Since 1800* (New York: McGraw-Hill, 1964), esp. Ch. 3, pp. 74-130.
18. L. M. Goreux, "Les Migrations Agricoles in France Depuis un Siecle et leur Relation avec Certains Facteurs Economiques," *Etudes et Conjuncture* (April 1956), 331.



F. Capital Migration and Interregional Linkages

Not only should the economy tend to develop a national labor market after experiencing continued secular growth, but more efficient national capital markets should evolve apace. If indeed perverse interregional capital flows had been typical of early growth stages, the development of more sophisticated capital markets in the Southern regions themselves should help deter the net outflow of capital. External economies and benefits accruing from agglomeration of capital projects may eventually become exhausted at the margin in the North while they begin to assert themselves in the poorer Southern regions as industrialization proceeds there (albeit, perhaps, at a slower rate). Finally, if growth becomes relatively rapid in the South due to any other factors, the capital flow will most likely undergo a natural reversal.<sup>19</sup> In Myrdal's terminology, the spread effects may begin to assert themselves from those islands of industrial growth as the economy fully integrates itself and commodity and factor markets become more efficient.

G. Central Government Policy

Perhaps most important, central governments may allow themselves the luxury of equality in the geographic distribution of income and pursue an active policy of income transfer to the poor regions. This may take the more dramatic form of TVA's, or regional concern may be implemented through highly-publicized institutions like the *Casa per il Mezzogiorno*, or it simply may result from a more general national commitment, not necessarily spatial, to welfare and equity. In the latter case, the appearance of a progressive income tax structure and concomitant welfare payments may be sufficient to create large regional transfers to the South without the emphasis on federal social overhead investment in the backward regions.

Finally, with regard to the central government's pattern of regional investment, it should be clear that after development has proceeded for some time, the need for public investment relative to private may tend to diminish, and in any case a larger portion of public investment may be financed from earnings of previous investments. This, of course, provides an excellent opportunity to alter the geographic composition of public investment in favor of the less developed areas.<sup>20</sup>

\* \* \*

Any one of these factors, or any combination of them, may be enough to cause regional inequality to diminish. Once the process of regional convergence or depolarization begins, however, it is likely to become cumulative, with the forces tending towards regional equality mutually strengthening each other contributing to a more rapid speed of adjustment.

The initial hypothesis of this study is, therefore, that the early stages of national development generate increasingly large North-South income differentials. Somewhere during the course of development, some or all of the disequilibrating tendencies diminish, causing a reversal in the pattern of interregional inequality. Instead of divergence in interregional levels of development, convergence becomes the rule, with the backward regions closing the development gap between themselves and the already industrialized areas. The expected result is

19. We certainly know very little about interregional capital flows, but Professor Lance Davis of Purdue is currently adding a great deal to our knowledge. In his study of American history, he thus far has found significant evidence of sharp reductions in interregional interest rate differentials from 1870 to 1914.

20. Hirschman, *op. cit.*, p. 194.

that a statistic describing regional inequality will trace out an inverted "U" over the national growth path; the historical timing of the peak level of spatial income differentials is left somewhat vague and may vary considerably with the resource endowment and institutional environment of each developing nation.

The rest of this paper summarizes the empirical evidence concerning the relation between levels of development and regional inequality.<sup>21</sup> To achieve this end and to utilize such data as exist, we have used alternative techniques. First, an international cross-section analysis is pursued for twenty-four countries during the decade of the 1950's. Second, the cross-section approach is applied to the United States census data (1950 and 1960) where counties are treated as the regional unit and the states as nations. Third, national time series analysis is applied to those few countries for which data is available.

Finally, we shall also attempt to shed light on three other related questions. (1) What is the relative importance of income growth versus population redistribution in contributing to the time pattern of regional inequality? (2) What role does the labor participation rate play in producing differences in income per capita levels? (3) Does regional inequality differ sharply between industrial sectors?

### III. International Cross-Section Analysis

The ideal measure of regional development in a study of this sort would be real income per capita (including income in kind) by geographic units which have maximum regional homogeneity. This ideal statistic is rarely available. First, the regional units are more or less given by the nature of decentralized political administrative units: for the United States the units are states; for Puerto Rico, *municipios*; for Canada, provinces; for Colombia, departments; for Spain, *provincias*. The regional units are not necessarily those which would be chosen by an economist or an economic geographer. Second, proper regional cost of living indices do not exist, and therefore any differential in regional price levels could possibly bias our results, since the data are usually given in the form of income in prices prevailing for the national unit. The nature of the regional units is unlikely to impart a systematic bias into the study, but regional price level differentials may be a serious problem, since the cost of living is usually lower in the poor agricultural regions. Furthermore, the divergence between regional price levels is likely to diminish as the nation develops.<sup>22</sup> Third, those regions which are primarily agricultural and which have less developed money economies will absorb a systematic downward bias, since their estimates of income rarely accurately record income in kind. The nature of the bias may vary with the level of national development as the Southern regions

21. If it has not been made so already, we should make it clear that this study does not concern itself with patterns of regional concentration of income and population over the national development spectrum. Our concern will be with the regional dispersion of per capita income and labor force productivity. It should be noted that the two concepts of regional concentration and regional income per capita differentials need not converge. For example, in the case of 20th century France, it appears that concentration of industry, income, and population around Metropolitan Paris and surrounding areas has been consistent with *convergence* in regional income per capita levels. On the other hand, with the United States, "it is interesting to observe that the lower rate of spatial redistribution of various countrywide aggregates toward the second half of the period [1900/10-1960] is accompanied by *greater* reduction in inequality of income per capita among regions." Simon Kuznets, *Population Redistribution and Economic Growth: United States, 1870-1950* (Philadelphia, 1960), pp. 270-71. It would be extremely fruitful to examine this aspect of the problem more intensively.
22. It is interesting to note that for one country where allegedly adequate regional cost of living indices are available, Finland (1950), the use of those price indices produced little effect upon our estimates of regional inequality.

also become fully monetized and market oriented. Finally, the income accounting concept (not to mention serious reservations about the reliability of the data themselves!) varies considerably from country to country.<sup>23</sup> Puerto Rican regional development levels are measured by median income per family, Norwegian by assessed income per capita, Canadian by personal income per capita, German by net national product per capita, and so on. It can only be hoped that none of these limitations is serious enough to negate the striking patterns discovered in the data.

Table I presents the results of the international cross-section study. Statistics for these twenty-four nations were available to us, and they are grouped according to Kuznets' seven level-of-development classifications. These twenty-four include thirteen European, four "empty" overseas European, four Latin American, and three Asian nations. Regression analysis was not attempted for this portion of the study because of the difficulty of cardinal ranking of these countries by levels of development or income per capita. Column 2 indicates the years from which the measures of inequality were computed. The period covered over-all ranges from 1949 to 1961. Where data for a number of years were available (as in the case of Italy, Norway, and the United States, for example), they were utilized to more closely approximate an average decade estimate of regional inequality.

Columns 3 and 4 give a measure of the extent of the "North-South problem" within these nations at widely differing levels of development. Column 3,  $V_w$ , is a weighted coefficient of variation which measures the dispersion of the regional income per capita levels relative to the national average while each regional deviation is weighted by its share in the national population;<sup>24</sup> the higher the  $V_w$ , the greater the size of geographic income differentials. Column 4,  $V_{uw}$ , is much less useful for our purposes since it is unweighted and will be determined in part by the somewhat arbitrary political definition of regional units (the number of which varies considerably between countries: see footnote to Table I). As a final word of caution preparatory to an examination of the results themselves, it should be noted that twenty of the twenty-four observations fall within groups I and IV or between "middle" and "high" income levels. This is indeed unfortunate, since it will not allow us to test significantly the hypothesis that  $V_w$  should rise, or in other words, that regional inequality should increase, during early stages of development. Furthermore, the sample does not include any of the communist East European nations, other than Yugoslavia, and this is a lamentable exclusion.

23. This criticism does not hold for the United States county-state study or for the time series studies which follow this section. In the case of the time series studies, however, the number and nature of the regional units sometimes vary over time.

24. More precisely,

$$V_w = \frac{\sqrt{\sum_i \frac{f_i}{n} (y_i - \bar{y})^2}}{\bar{y}},$$

where  $f_i$  = population of the  $i^{\text{th}}$  region,

$n$  = national population,

$y_i$  = "income per capita" of the  $i^{\text{th}}$  region,

$\bar{y}$  = national income per capita,

$$\text{and } V_{uw} = \frac{\sqrt{\sum_i \frac{(y_i - \bar{y})^2}{N}}}{\bar{y}}, \quad \text{where } N = \text{number of regions.}$$

As a brief reading of the footnotes to Table I will suggest, this study utilizes a more detailed regional breakdown than usually appears in the literature. For example, we computed  $V_w$  from nineteen Italian regions rather than unnecessarily limit ourselves to the conventional separation into North, Central, and South.

TABLE 1.

## International Cross-Section

| Country and Kuznets<br>group classification<br>(1) | Years covered<br>(2) | $V_w$<br>(3) | $V_{uw}$<br>(4) | $M_w$<br>(5) | Size<br>(square miles)<br>(6) |
|--|----------------------|--------------|-----------------|--------------|-------------------------------|
| Australia  | 1949/50-1959/60      | .058         | .078            | 4.77         | 2,974,581                     |
| New Zealand  | 1955                 | .063         | .082            | 4.93         | 103,736                       |
| Canada   | 1950-61              | .192         | .259            | 17.30        | 3,845,774                     |
| United Kingdom                                     | 1959/60              | .141         | .156            | 11.39        | 94,279                        |
| United States                                      | 1950-61              | .182         | .189            | 16.56        | 3,022,387                     |
| Sweden   | 1950, '55, '61       | .200         | .168            | 15.52        | 173,378                       |
| Group I average                                    |                      | .139         | .155            | 11.72        |                               |
| Finland  | 1950, '54, '58       | .331         | .276            | 26.64        | 130,165                       |
| France   | 1954, '55/56, '58    | .283         | .215            | 20.80        | 212,659                       |
| West Germany                                       | 1950-55, '60         | .205         | .205            | 16.98        | 94,723                        |
| Netherlands  | 1950, '55, '58       | .131         | .128            | 12.45        | 12,850                        |
| Norway   | 1952, '57-'60        | .309         | .253            | 23.84        | 125,064                       |
| Group II average                                   |                      | .252         | .215            | 20.14        |                               |
| Ireland  | 1960                 | .268         | .271            | 24.20        | 26,601                        |
| Chile  | 1958                 | .327         | .440            | 30.65        | 286,397                       |
| Austria  | 1957                 | .225         | .201            | 18.69        | 32,369                        |
| Puerto Rico  | 1960                 | .520         | .378            | 42.31        | 3,435                         |
| Group III average                                  |                      | .335         | .323            | 28.96        |                               |
| Brazil   | 1950-59              | .700         | .654            | 53.78        | 3,288,050                     |
| Italy  | 1951, '55, '60       | .360         | .367            | 30.94        | 117,471                       |
| Spain  | 1955, '57            | .415         | .356            | 32.32        | 195,504                       |
| Colombia   | 1953                 | .541         | .561            | 46.70        | 439,617                       |
| Greece   | 1954                 | .302         | .295            | 26.56        | 51,246                        |
| Group IV average                                   |                      | .464         | .447            | 38.06        |                               |
| Yugoslavia   | 1956, '59, '60       | .340         | .444            | 24.54        | 95,558                        |
| Japan  | 1951-59              | .244         | .222            | 19.98        | 142,644                       |
| Group V average                                    |                      | .292         | .333            | 22.26        |                               |
| Philippines  | 1957                 | .556         | .627            | 29.59        | 115,600                       |
| Group VI average                                   |                      | .556         | .627            | 29.59        |                               |
| India  | 1950/51, 1955/56     | .275         | .580            | 19.39        | 1,221,880                     |
| Group VII average                                  |                      | .275         | .580            | 19.39        |                               |
| Total average                                      |                      | .299         | .309            | 23.78        |                               |

Table 1 (continued)

Definitions (see Appendix Tables for source descriptions and for a more extensive description of regional units):

- (1) Australia, 1949/50-1959/60. Based on personal income per capita. Six regions: New South Wales (including Australian Capital Territory), Victoria, Queensland, South Australia (including Northern Territory), Western Australia, and Tasmania.
- (2) New Zealand, 1955. Based on personal income per capita. Ten provincial districts and sectors: Auckland, Hawkes Bay, Taranaki, Wellington, Marlborough, Nelson, Westland, Canterbury, Otago, and Southland.
- (3) Canada, 1950-1961. Based on personal income per capita. Eleven provinces: Newfoundland, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia, and Yukon and Northwest Territories.
- (4) United Kingdom, 1959/60. Based on total net assessed income per capita. Fifteen regions: nine English "Standard Regions," West Central Scotland, East Central Scotland, Highlands, Scottish Border Counties, Northern Ireland, and Wales.
- (5) United States, 1950-61. Based on personal income per capita. 1950-54  $V_w$  estimates are from Frank A. Hanna, *State Income Differentials, 1919-1954* (Durham: Duke University Press, 1959), p. 36, and  $V_{uw}$  was derived from Hanna's data (T-4, pp. 38-41); 1955-61, income per capita estimates are from various issues of the *Survey of Current Business*, and population estimates are taken from the *Statistical Abstract for the United States*. Nine regions based upon Bureau of Census groupings: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific.
- (6) Sweden, 1951, 1955, 1961. Based on assessed income per capita. Twenty-four *lans* plus Stockholms *stad*: Stockholms *stad*, Stockholms *lan*, Uppsala, Södermanlands, Östergötlands, Jönköpings, Kronobergs, Kalmar, Gotlands, Blekinge, Kristianstads, Malmöhus, Hallands, Göteborgs o. Boh., Älvsborgs, Skaraborgs, Varmlands, Örebro, Västmanlands, Kopparbergs, Gävleborgs, Västernorrlands, Jämtlands, Västerbottens, and Norrbottens.
- (7) Finland, 1950, 1954, 1958. Based on declared income per capita. Twenty-three "economic regions" for 1950. Ten provinces for 1954, 1958: Uudenmaan, Turun-Porin, Ahvenanmaa, Hämeen, Kymen, Mikkelin, Kuopion, Vaasan, Oulun, and Lapin.
- (8) France, 1954, 1955/56, 1958. Based on disposable income per capita for 1954 and personal income per capita for 1955/56 and 1958. Twenty-one regions: Alsace, Aquitaine, Auvergne, Bourgogne, Bretagne, Centre, Champagne, Franche-Comté, Languedoc, Limousin, Lorraine, Midi-Pyrénées, Nord, Normandie (Basse), Normandie (Haute), Pays de la Loire, Picardie, Poitou-Charentes, Provence, Région parisienne, and Rhône-Alpes.
- (9) West Germany, 1950-1955, 1960. Based on net product at factor cost per capita. Nine provinces of West Germany (excluding the Saar and Berlin): Schleswig-Holstein, Hamburg, Niedersachsen, Bremen, North Rhein-Westfalen, Hessen, Rheinland-Pfalz, Baden-Württemberg, and Bayern.
- (10) Netherlands, 1950, 1955, 1958. Based on net product at factor cost per capita. Eleven provinces: Groningen, Friesland, Drenthe, Overijssel, Gelderland, Utrecht, Noord-Holland, Zuid-Holland, Zeeland, Noord-Brabant, and Limburg.
- (11) Norway, 1952, 1957-60. Based on assessed income per capita. Twenty counties: Østfold, Akershus, Oslo, Hedmark, Oppland, Buskerud, Vestfold, Telemark, Aust-Agder, Vest-Agder, Rogaland, Hordaland, Bergen, Sogn og Fjordane, Møre og Romsdal, Sør-Trøndelag, Nord-Trøndelag, Nordland, Troms, and Finnmark.
- (12) Ireland, 1960. Based on earned income per capita. Twenty-six counties: Carlow, Dublin, Kildare, Kilkenny, Laoighis, Longford, Louth, Meath, Offaly, Westmeath, Wexford, Wicklow, Clare, Cork, Kerry, Limerick, Tipperary, Waterford, Galway, Leitrim, Mayo, Roscommon, Sligo, Cavan, Donegal, and Monaghan.

Table 1 (continued)

- (13) Chile, 1958. Income concept not given in source. Nine regions: Tarapaca y Antofagasta, Atacama y Coquimbo, Aconcagua y Valparaiso, Santiago y O'Higgins, Colchagua y Curico y Talca y Mank y Linares, Nuble y Concepcion y Arauco y Bio-Bio, Malleco y Cautin, Valdivia y Osorno y Llanquihue y Chiloé y Aysen, and Magallanes.
- (14) Austria, 1957. Based on national income per capita. Nine provinces: Wien, Niederösterreich, Oberösterreich, Steiermark, Tirol, Kärnten, Salzburg, Vorarlberg, and Burgenland.
- (15) Puerto Rico, 1960. Based on median family income. Income and population estimates are from the 1960 United States *Census of Population*, T-35 and T-37, pp. 116 and 117. Seventy-six *municipios*.
- (16) Brazil, 1950-59. Based on national income per capita. Twenty-one states: Amazonas, Para, Maranhão, Piaui, Ceara, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia, Minas Gerais, Espirito Santo, Rio de Janeiro, Est. du Guanabara, São Paulo, Parana, Santa Catarina, Rio Grande do Sul, Mato Grosso, and Goias.
- (17) Italy, 1951, 1955, 1960. Based on net national product per capita. Nineteen regions: Piemonte, Valle d'Aosta, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia G., Liguria, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Abruzzi e Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, and Sardegna.
- (18) Spain, 1955, 1957. Based on national income per capita. Fifty provinces.
- (19) Colombia, 1953. Income concept not clear in source: given as "income" per capita. Sixteen departments: Antioquia, Atlantico, Bolivar, Boyaca, Caldas, Cauca, Cordoba, Caudinamarca, Choco, Huila, Magdalena, Nariño, Norte de Santander, Santander, Tolima, and Valle.
- (20) Greece, 1954. Based on national income per capita. Eleven regions: Sterea Hellas and Eubolia, Macedonia, Aegean Islands, Peloponnesos, Cyclades, Thessaly, Crete, Dodecanesos, Thrace, Ionian Islands, and Epirus.
- (21) Yugoslavia, 1956, 1960. Based on national income per capita. Six provinces in 1956: Serbia, Croatia, Slovenia, Bosnia and Hercegovina, Macedonia, and Montenegro. Eight provinces in 1960: Serbia is broken down into the sub-regions of Serbia Proper, Vojvodina, and Kosovo and Metohiya.
- (22) Japan, 1951-59. Based on personal income per capita. Forty-six prefectures.
- (23) Philippines, 1957. Based on personal income per family. Ten regions: Metropolitan Manila, Ilocos and Mt. Province, Cagayan Valley and Batanes, Central Luzon, Southern Luzon and Marinduque and Mindoro and Palawan, Bicol Province, Western Visayas, Eastern Visayas, South West Mindanao and Sulu, and North East Mindanao.
- (24) India, 1950/51, 1955/56. Based on national income per capita. Eighteen states: Andhra, Assam, Bihar, Gujarat, Kerela, Madhya Pradesh, Madras, Maharashtra, Mysore, Orissa, Punjab, Rajasthan, Uttar Pradesh, West Bengal, Delhi, Himachal Pradesh, Manipur, and Tripura.

The results are quite striking. Our measure of regional inequality,  $V_w$ , ranges widely between a maximum of 0.700 for Brazil, a country with the most widely publicized North-South problem, and a minimum of 0.058 for Australia. If we examine the averages for each income class, the relationship between level of development and degree of regional inequality appears to be quite significant. The severity of the North-South problem seems to be quite minor indeed among the mature economies, although for any of these countries it may be politically significant. Group I has an average  $V_w$  of 0.139; but the degree of inequality increases sharply as we move from the high income to middle income group where the index of regional inequality measures 0.464, between three or four times that of the high income group. It is interesting to note that the middle income group contains both of those nations which are universally noted for their severe North-South schism, Brazil with its Nordeste, and Italy with its Mezzogiorno. As we move from Group IV to VII the evidence becomes thin, but we have only one exception, the Philippines, to the hypothesis that regional inequality should increase

in the early stages of growth. Both groups V and VII have average  $V_w$ 's significantly less than the middle income class, and India appears to have the less serious regional income problem than the average of Yugoslavia and Japan. With one exception, it does appear that the pattern of regional inequality is in the form of an inverted "U," reaching a peak in the middle income class.<sup>25</sup> However, it should be pointed out that the evidence supporting a tendency towards increasing regional divergence from group VII to group IV is of a much weaker sort than that supporting the convergence stage from group IV to group I. Of the four observations falling between the low and middle income classes, three support the hypothesis, and two of these, Japan and India, may be unusually low relative to their income class averages. That is, Kuznets' indices of inequality based on industrial sectors reveal unusually low measures for both those nations relative to their respective income classes.<sup>26</sup>

The variation within each of these income classes is in some cases quite large and extremely interesting as well. Among the most developed nations, Canada, the United States, and Sweden, all have significantly more serious North-South problems than the average for income class I as a whole. With the exception of Australia, these three nations also have the largest land mass. The suggestion here is that geographic size may secondarily influence the degree of regional inequality; given the level of national development, the larger the geographic size of the national unit, the greater will be the degree of regional inequality. This could be explained by any number of factors that may already appear obvious to the reader; the greater the geographic size, the larger the scope for wide regional variations in natural resource endowment due to increased distance, both economic and cultural; the weaker the linkages between regions and the stronger the incidence of localism. The relatively low  $V_w$ 's in New Zealand and the United Kingdom are both consistent with this notion, since their national borders encompass small geographic areas. Australia would appear to be an exception to this generalization, but even this conflict, large land mass but a minimum problem with regional inequality, can be explained by the arid nature of most of that continent; only the coastal areas are densely populated.

The same relationship appears to hold for less developed nations as well. In group II, both West Germany and the Netherlands have relatively low degrees of regional inequality for the postwar period, and both are significantly smaller in geographic size than the rest of the group. In group III, however, tiny Puerto Rico is a glaring exception to this generalization, although again Austria and Ireland with small land areas have low  $V_w$ , while Chile reveals a high incidence of regional dualism consistent with its large size. The evidence is equally strong among the middle income nations of group IV: Brazil and Colombia are both large relative to the rest of that group, and both have North-South problems more severe than that of the rest of the group. Greece, with a relatively low  $V_w$ , is less than half as large as the next biggest nation, Italy. No attempt was made to pursue this relationship further in groups V-VII, since the number of observations is obviously too small.<sup>27</sup>

25. Using available national income per capita estimates [Hollis B. Chenery, "Patterns of Industrial Growth," *American Economic Review*, L (September 1960), Table 1, 632], the Spearman rank correlation coefficient is 0.721 for the nineteen nations in groups I-IV. These results are consistent with those of the only other cross-section study dealing with regional inequality with which we are familiar. The *Economic Survey of Europe in 1954* revealed an association between high national income per capita and low degrees of regional inequality (pp. 136-71).

26. Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations. VIII. Distribution of Income by Size," *Economic Development and Cultural Change*, XI, No. 2, Part II (January 1963), Appendix Table I, 70-71.

27. It should be clear that our "size" variable is very crude as a measure of interregional migration barriers and incidence of non-homogeneity within a nation. The Philippines is an excellent example of a country of moderate size but with tremendous natural barriers to migration; the nature of its geography



## REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 171

Although we have no way of evaluating the political importance of regional equality of income distribution to any one nation, it does seem strange that so many of these countries in our sample feel that their North-South problems are especially unique and severe. For instance, the French concern with "Paris and the French desert"<sup>28</sup> seems somewhat extravagant, given the size and level of development of that nation.<sup>29</sup> For that matter, Italy's highly popularized schism between North and South does not appear to be a severe case of regional dualism at all, given its income level. It should be made clear again, however, that we have made no attempt to measure regional concentration of population and urbanization, but only have measured regional variation in per capita income levels weighted by the distribution of population.

One more comment might be made before examining the *changes* in regional inequality during the postwar era. Recall that our weighted index of regional inequality,  $V_w$ , involves the squaring of the per capita income differentials. Although this is useful for the analysis of variance which appears in Section VIII, nevertheless it is conceivable that by squaring the differences we may be making our index unnecessarily sensitive to a few extreme deviations in regional per capita income. In order to check our results, we used an alternative statistic which sums the differentials to the first power signs disregarded.<sup>30</sup> This alternative measure,  $M_w$ , appears in column 5:  $M_w$  appears to produce significantly different results only in the case of the Philippines, where our alternative index of regional inequality is much more consistent with that country's level of development (see Table 1).

Table 2 presents us with some more information regarding the problem of regional inequality. There we have summarized the recent *changes* in geographic income differentials for those countries for which such short-term time series are available. Sixteen of the original twenty-four nations are classified according to the direction of change in  $V_w$  since the mid-late 1940's. What we hoped to find here were movements consistent with each country's position on the development spectrum. Divergence of regional income per capita levels should generally hold true for those countries below the middle income group, while convergence should be the case for those above the middle income range. That is, India at low levels of per capita income and regional disparity should exhibit increasing regional dualism and a rising  $V_w$  as

may help explain its apparent high degree of regional inequality. Furthermore, severe regional dualism may in part be explained by historical accident. Yugoslavia's unusual history, which has produced such a high degree of ethnic, religious, and linguistic non-homogeneity, surely explains a large part of their contemporary problems with regional inequality.

28. This is J-F. Gravier's term. See his *Paris et le desert francais* (Paris: Le Portulan, 1947).
29. It should be emphasized again, however, that we are not referring to *absolute* regional differentials but relatives: sharp declines in percentage regional deviations from the national average may be quite consistent with increasing absolute gaps in regional income levels. Our own feeling is that the latter measure is not very helpful in understanding the inequality process; but, as Professor Benjamin Higgins has pointed out privately, if "the policy issue is how to reduce the gap in productivity and income between leading and lagging regions, the absolute gap between richest and poorest regions in the country may be more significant than [an] index of dispersion."

30. In this case,

$$M_w = \frac{\sum |y_i - \bar{y}| \frac{f_i}{n}}{\bar{y}} \times 100$$

where  $f_i$  = population of the  $i^{\text{th}}$  region,  
 $n$  = national population,  
 $y_i$  = "income per capita" of the  $i^{\text{th}}$  region,  
 $\bar{y}$  = national income per capita.



TABLE 2  
Secular Changes in  $V_w$  During The Postwar Period

| Income<br>class | $V_w$<br>rising     | $V_w$<br>stable             | $V_w$<br>falling                                 |
|-----------------|---------------------|-----------------------------|--|
| I               |                     | Australia<br>United Kingdom | Canada<br>United States<br>Sweden                |
| II              |                     | France                      | Finland<br>West Germany<br>Netherlands<br>Norway |
| III             |                     |                             |  |
| IV              |                     | Italy                       | Spain<br>Brazil                                  |
| V               | Japan<br>Yugoslavia |                             |  |
| VI              |                     |                             |  |
| VII             | India               |                             |  |

Source: See Appendix and Table 4.

she proceeds through her early stages of modern development; the mature U.S. economy at high per capita income levels should be undergoing a further diminution in what is already a low degree of regional inequality and imbalance.

And indeed this is most strikingly the case: the nations in groups I and II exhibit either stability or a weakening in their North-South problems. Seven of the ten nations in these two groups underwent a decline in  $V_w$ . Those in the middle income group show a variety of change consistent with the pivotal nature of that level of development: Brazil underwent a decline from a "secular peak" in regional inequality, Italy exhibited stability in  $V_w$  at a high level, and Spain experienced regional convergence.<sup>31</sup> The remaining nations of the low income groups all reveal tendencies towards divergence in regional income levels.

We should note, at this stage of our discussion, that there is a significant amount of information for the currently underdeveloped nations which this study has not employed (either due to its unreliability or due to its non-quantitative nature). For example, there does exist some information which indicates the increasing problem of regional divergence in Pakistan's

31. Lasuen implies that Spanish experience has been just the opposite. He suggests that regional divergence has been the case for the past decade. *Op. cit.*, pp. 169-88.

postwar dev  
but they do

1951-52  
1952-53  
1953-54  
1954-55  
1955-56  
1956-57  
1957-58  
1958-59  
1959-60

At any  
international

If we l  
countries as  
Using the m  
States censu  
in each state  
of advantage  
sample size  
Second, the  
data as well  
capita and if  
differ signifi  
reason, any  
tials is more  
regards to s

12. S. U. K  
ment Re  
distribut

13. This doe  
intercou  
tivity, a  
via cent  
regional  
their res  
gional b  
expende  
stimulat

## REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 173

postwar development. The figures below are meant only to be suggestive, not conclusive, but they do stress regional divergence in growth rates between East and West Pakistan.<sup>12</sup>

Pakistan (per capita quantum indices)

|         | West Pakistan | East Pakistan | Pakistan |
|---------|---------------|---------------|----------|
| 1951-52 | 100           | 100           | 100      |
| 1952-53 | 101           | 103           | 102      |
| 1953-54 | 107           | 115           | 111      |
| 1954-55 | 115           | 105           | 110      |
| 1955-56 | 114           | 91            | 103      |
| 1956-57 | 118           | 116           | 117      |
| 1957-58 | 124           | 109           | 117      |
| 1958-59 | 125           | 96            | 111      |
| 1959-60 | 128           | 112           | 120      |

At any rate, these short-term time series movements appear to be consistent with our international cross-section results.

### IV United States Cross-Section Analysis, 1939 and 1960

If we treat the states within American borders as nations themselves, and define counties as the regional unit, we ought to be able to perform an independent cross-section test. Using the median family income and population estimates which are enumerated in the United States census by county, we can then determine the degree of regional inequality that exists in each state at varying levels of per capita income and development. This test has a number of advantages over the international cross-section analysis summarized above. First, the sample size is twice as large, since it includes forty-six states (see footnote to Table 1). Second, the sample has the advantage of utilizing both more reliable income and population data as well as more comparable income data, making a cardinal ranking of state income per capita and the use of regression analysis less objectionable. Third, the states themselves differ significantly from nations in terms of their control over economic activity. For this reason, any relationship between levels of state development and intercounty income differentials is more likely to be attributable to "natural" forces rather than governmental policy with regards to spatial inequality.<sup>13</sup>

12. S. U. Khan, "A Message of Economic Growth in East and in West Pakistan," *The Pakistan Development Review*, I, No. 2 (Autumn 1961), 31, Table II, and 39, Table I. Furthermore, population redistribution has been in favor of West Pakistan during this decade.

13. This does not mean that governmental policy cannot appear as an important explanatory variable of intercounty income differentials. Each state does have some autonomy of control over economic activity, and, furthermore, the national government may itself influence intercounty income differentials via central government policies. As an example of the impact which state government may have upon regional inequality, many of the currently industrializing Southern states are attempting to diversify their new industry spatially. The Mississippi Agricultural and Industrial Board has recently given regional balance as one of its primary goals by bringing industry to the people. That is, they have expressed an interest in spreading industrial projects evenly throughout the state, rather than further stimulating the flow of interstate migration into the major urbanized counties.

At the same time, however, the data presents us with significant disadvantages. First, we have already mentioned that county income data is expressed, approximately, in terms of personal income (rather than "national" income) and per family (rather than per capita). The use of median income *per family* may introduce a significant bias if there are substantial state differences in rural-urban family size differentials. Thus, there *may* be a systematic bias which tends to minimize the observed size of intercounty income differences in the poor states. The second disadvantage is that the income range from poorest to richest state is much more narrow than that for our international cross-section. In particular, the U.S.'s poorest state, Mississippi, is significantly above the Kuznets middle income range, and for that reason alone we cannot expect to find evidence which would shed light on that part of the hypothesis which predicts rising inequality during early stages of economic growth. What we should find is that the more developed states in the Northeast and Midwest have very small intercounty differences in income levels relative to the lower income states.

The results of these tests for both 1950 and 1960 seem to throw added support behind our contentions concerning regional inequality. Table 3a summarizes the results by listing the computed regional inequality measures by state (where  $V_w$  is determined in precisely the same fashion as in Section III). For 1950, the range lies between a low degree of regional inequality for Connecticut, 0.0627, and a high  $V_w$  for Georgia, 0.3965. Table 3b aggregates this evidence into a summary table where the inverse relationship between levels of state development and regional inequality seems fairly clear. On the average, the eight lowest income states have a coefficient of intercounty inequality approximately two and one-half times that of the richest seven. The same pattern holds true for the 1960 data, where again severe interregional differentials are associated with relatively low levels of development.

Using simple univariate regression analysis, the inverse correlation between state income per capita and our index of interregional inequality is significant for both years:

$$\begin{array}{lll} [1950] & V_w = 0.52792 - 0.000131 Y_{pc}50, & R = 0.760 \\ & (0.000017) & \\ [1960] & V_w = 0.46791 - 0.000139 Y_{pc}60, & R = 0.687 \\ & (0.000022) & \end{array}$$

This presents us with a second question: What role does state size play in determining the degree of regional inequality? At first blush, the results were definitely negative, and geographic size did not appear to be a significant determinant of intercounty income differentials. Geographic size is either a poor proxy variable for measuring, say, diversity of state resources, or the importance of that independent variable in the international cross-section simply does not appear in a comparative analysis among American states. It is the former which seems to explain our poor results, since when we exclude the Mountain and Pacific states plus Texas, geographic size becomes a significant determinant of  $V_w$ .<sup>34</sup> The justification for this exclusion is approximately the same as our treatment of Australia in the international cross-section. For most, but certainly not all (e.g., California), of these states, large geographic size is not identical with varied resources or significant localism but with empty regions of semi-desert.

At any rate, for the states east of the Mississippi and on the Great Plains, size is a significant determinant of regional inequality. The higher the level of development and the

34. The states excluded here are Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Texas, Utah, Washington, and Wyoming.

# REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 175

smaller the state, the lower is the index of regional inequality:<sup>35</sup>

$$[1950] \quad V_w = 0.4384 - 0.000113 Y_{pc}^{50} + 0.001205 S, \quad R = 0.791$$

(0.000019) (0.000461)

One more question remains to be answered. Has the pattern of change in intrastate inequality been consistent with each state's position on the development spectrum? Furthermore, is it consistent with the postwar experience of the United States with *interstate* inequality trends? Both of these questions may be answered in the affirmative. Column 4 of Table 3a exhibits the change in  $V_w$  for each state between the terminal years of the decade 1950-60. In only nine of the forty-six states did the size of regional inequality increase, and there appears to be no common factor among them: Iowa, Minnesota, Nebraska, North Dakota, Texas, and West Virginia are the major exceptions, while Massachusetts, North Carolina, and Pennsylvania exhibit minor increments in  $V_w$  over the decade. During a decade when regional inequality *among* the states was declining sharply (see Section V), regional inequality *within* the majority of the states was declining as well. And to repeat, these postwar movements are consistent with the fact that all of the U.S. states are above the middle income range and in relatively mature states of economic growth.

TABLE 3a.

United States Cross-Section, 1950 and 1960

| State       | $V_w$<br>(1950) | $V_w$<br>(1960) | $\Delta V_w$ | $Y_{pc}$<br>(1950) | Geographic size<br>(land area:<br>1,000 sq. miles) |
|-------------|-----------------|-----------------|--------------|--------------------|--|
| (1)         | (2)             | (3)             | (4)          | (5)                | (6)  |
| Alabama     | 0.3529          | 0.280           | -0.0729      | 1956               | 51.1   |
| Arizona     | 0.1639          | 0.112           | -0.0519      | 2375               | 113.6  |
| Arkansas    | 0.3356          | 0.292           | -0.0436      | 1315               | 52.7   |
| California  | 0.1045          | 0.099           | -0.0055      | 3021               | 156.7  |
| Colorado    | 0.1659          | 0.163           | -0.0029      | 2514               | 103.9  |
| Connecticut | 0.0627          | 0.053           | -0.0097      | 3155               | 4.9  |
| Florida     | 0.2171          | 0.147           | -0.0701      | 1950               | 54.3   |
| Georgia     | 0.3965          | 0.300           | -0.0965      | 1649               | 58.5   |
| Idaho       | 0.1378          | 0.121           | -0.0168      | 2685               | 82.8   |
| Illinois    | 0.1686          | 0.167           | -0.0016      | 3163               | 55.9   |
| Indiana     | 0.2005          | 0.136           | -0.0645      | 2827               | 36.2   |
| Iowa        | 0.1663          | 0.201           | +0.0347      | 2612               | 56.0   |
| Kansas      | 0.2389          | 0.211           | -0.0279      | 2377               | 82.1   |
| Kentucky    | 0.3908          | 0.352           | -0.0388      | 1774               | 39.9   |
| Louisiana   | 0.2916          | 0.267           | -0.0246      | 1810               | 45.2   |
| Maine       | 0.1269          | 0.110           | -0.0169      | 2213               | 31.0   |
| Maryland    | 0.2483          | 0.223           | -0.0253      | 2811               | 9.9  |

35. It has been suggested that the number of regional units, counties, within each state may significantly affect our statistic of regional inequality. That is, the greater the number of counties within the state, everything else being equal, the larger the  $V_w$ . Given that state income per capita and the number of counties within the state are independent, then is it possible that our "size" variable does nothing more than reflect the impact of the number of counties upon state  $V_w$ ? Interestingly enough, however, state size and number of counties are also independent of each other (Spearman rank correlation coefficient is 0.077).

Table 3a (continued)

| (1)            | (2)    | (3)   | (4)     | (5)  | (6)   |
|----------------|--------|-------|---------|------|-------|
| Massachusetts  | 0.0854 | 0.092 | +0.0066 | 2909 | 7.9   |
| Michigan       | 0.1791 | 0.124 | -0.0551 | 3195 | 57.0  |
| Minnesota      | 0.1980 | 0.236 | +0.0380 | 2683 | 80.0  |
| Mississippi    | 0.3862 | 0.366 | -0.0202 | 1028 | 47.2  |
| Missouri       | 0.3622 | 0.301 | -0.0612 | 2200 | 69.2  |
| Montana        | 0.1686 | 0.146 | -0.0226 | 2718 | 145.9 |
| Nebraska       | 0.1617 | 0.238 | +0.0763 | 2389 | 76.7  |
| Nevada         | 0.1243 | 0.094 | -0.0303 | 2982 | 109.8 |
| New Hampshire  | 0.1067 | 0.056 | -0.0507 | 2405 | 9.0   |
| New Jersey     | 0.1440 | 0.110 | -0.0340 | 3285 | 7.5   |
| New Mexico     | 0.3293 | 0.227 | -0.1023 | 2301 | 121.5 |
| New York       | 0.1739 | 0.152 | -0.0219 | 3055 | 47.9  |
| North Carolina | 0.2696 | 0.274 | +0.0056 | 1864 | 49.1  |
| North Dakota   | 0.1461 | 0.204 | +0.0579 | 2446 | 70.1  |
| Ohio           | 0.1599 | 0.120 | -0.0399 | 3024 | 41.0  |
| Oklahoma       | 0.3129 | 0.252 | -0.0609 | 2050 | 69.0  |
| Oregon         | 0.0921 | 0.077 | -0.0151 | 2933 | 96.3  |
| Pennsylvania   | 0.1339 | 0.138 | +0.0041 | 2834 | 45.0  |
| Rhode Island   | 0.1066 | 0.050 | -0.0566 | 2650 | 1.1   |
| South Carolina | 0.3102 | 0.229 | -0.0812 | 1647 | 30.3  |
| South Dakota   | 0.3037 | 0.252 | -0.0517 | 2337 | 76.5  |
| Tennessee      | 0.3160 | 0.288 | -0.0280 | 1749 | 41.8  |
| Texas          | 0.1755 | 0.242 | +0.0665 | 2273 | 263.5 |
| Utah           | 0.1443 | 0.109 | -0.0353 | 3001 | 82.3  |
| Vermont        | 0.1135 | 0.094 | -0.0195 | 2101 | 9.3   |
| Washington     | 0.1351 | 0.112 | -0.0231 | 2955 | 66.8  |
| West Virginia  | 0.2181 | 0.230 | +0.0119 | 2344 | 24.1  |
| Wisconsin      | 0.2102 | 0.183 | -0.0272 | 2860 | 54.7  |
| Wyoming        | 0.1384 | 0.115 | -0.0234 | 2964 | 97.5  |

TABLE 3b.

| Income groups | Average $V_w$<br>(1950) | Average $V_w$<br>(1960) | No. of states<br>in group |
|---------------|-------------------------|-------------------------|---------------------------|
| I             | 0.1367                  | 0.114                   | 7                         |
| II            | 0.1510                  | 0.123                   | 8                         |
| III           | 0.2000                  | 0.179                   | 8                         |
| IV            | 0.1585                  | 0.159                   | 7                         |
| V             | 0.2616                  | 0.217                   | 8                         |
| VI            | 0.3134                  | 0.286                   | 8                         |

**Source:** Of the continental states, this group does not include Delaware, due to the small number of counties in that state (three), or Virginia, due to a unique method of regional classification (into urban units). The 1950 and 1960 median income per family and population data for counties are from Tables 36 and 35 in the 1950 and 1960 *United States Census of Population*, United States Department of Commerce, Bureau of the Census, "Detailed Characteristics by State." In some of these states a few counties are eliminated from the computations, since the Census does not estimate median family income if the population size is below a low minimum.

Table 3b (c)

Th  
"Que  
by Si  
75-7  
Th  
foun

The c  
consistent  
underwent t  
severe Nort  
their advan  
trace out of  
parent nati  
appear to b  
such rand

The p  
cant of whi  
nation back  
regional poi  
Although the  
data for onl  
back for mo  
nations are  
(1840-1961)  
lands (1938  
Germany (1  
World War I  
Germany co  
dualism (the  
possibilitie

The it  
regional ine  
exceptions  
reduced regi

36. Commer  
structur  
where t  
but it is  
income  
have ha  
such as  
have no  
Europea  
"Econoe

37. Italy an  
The Ger  
it in the

Table 3b (continued)

The cutoff points defining income groups in Table 3b are taken from Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations. VIII. Distribution of Income by Size," *Economic Development and Cultural Change*, XI, No. 2, Part II (January 1961), 75-77.

The state income per capita figures ( $Y_{pc}$ ) are official government estimates and can be found in the *Survey of Current Business*.

#### V The Historical Patterns: One

The question now arises as to whether the long-run historical experience of nations is consistent with our cross-section results. As the currently more mature national economies underwent the early process of economic development, did they experience first increasingly severe North-South dualism which eventually reached a peak and declined thereafter during their advanced stages of growth? Did the "empty" countries with ever-expanding frontiers trace out changing patterns of regional inequality quite different from those of the settled parent nations in Europe? <sup>36</sup> Does the experience of these nations with regional inequality appear to be one of a smooth trend, or is the degree of spatial inequality widely disturbed by such random factors as war, discovery, and political change?

The problems surrounding this methodological approach are immense, the most significant of which is the predictable lack of data necessary to extend a time series study for any nation back into its formative or adolescent years of economic development. Even where such regional population and income data are available, their reliability is usually very questionable. Although there may well be sources which we left untapped, we were able to find historical data for only ten nations which extended their experience with regional income differentials back for more than two decades. Our quantitative measures of regional inequality for these nations are presented in Table 4, and they cover the following periods: United States (1840-1961), United Kingdom (1937-59/60), France (1864-1958), Canada (1926-60), Netherlands (1938-58), Sweden (1920-61), Norway (1939-60), Italy (1928-60), Brazil (1939-59), and Germany (1900-60). Only six of these cover periods which extend back significantly before World War II, and of these six it would appear that only the United States, France, and Germany cover sufficiently long periods to encompass the three hypothesized stages of regional dualism (increasing, stable, and declining regional inequality), with Sweden and Italy dubious possibilities.

The first observation of interest might be the apparently consistent pattern of change in regional inequality between the 1930's and the immediate postwar period. With the predictable exceptions of Italy and Brazil,<sup>37</sup> each of these nations exhibit tremendous changes towards reduced regional disparity during that decade, all of them experienced significant convergence

36. Commenting on the hypothesized inverted "U" traced out by historical experience with income structure, Kuznets has said, "This long secular swing would be most pronounced for older countries where the dislocation effects of the earlier phases of modern economic growth were most conspicuous, but it might be found in the 'younger' countries like the United States...." With regard to spatial income distribution, this is apparently not the case for the "younger" Latin American nations. They have had historically more severe cases of regional dualism than that of the older European states such as Germany, France, and even Italy. However, it is clear that both Canada and the United States have not had as severe problems with regional income inequality as the Scandinavian and Continental European nations (excluding Germany: see immediately below and Sections III and VII). Simon Kuznets, "Economic Growth and Income Inequality," *American Economic Review*, XLV (March 1955), 1-28.

37. Italy and Brazil, of course, do not belong to the same high income groups as the other eight nations. The German series suffers from too much discontinuity between 1936 and 1939 to enable us to include it in the analysis which follows below.

in regional income levels. Surely a large part of the apparent lessening in regional inequality during this period can be explained by the unusually high levels of regional disparity reached by most nations during the 1930's. The "short-run" effects of the Great Depression were felt much more severely in the agricultural regions of each country, especially the United States and Canada, and a large part of the regional convergence in per capita incomes from the 1930's to the 1940's was due simply to a movement back to national full employment.<sup>38</sup> A good part of this convergence may also be explained by the war itself, which tended to strengthen secular forces towards convergence. How much of this apparent convergence is due to overt government policy and concern with regional inequality is debatable: no answer regarding causation is attempted at this stage.

Let us move on to the major question. What has been the long-term relationship between regional inequality and economic development?

In spite of the fact that the United States has undergone a growth experience somewhat unique among nations, being an "empty" country exhibiting unusual regional aspects in its development process, it traces out a "classic" pattern of regional inequality.<sup>39</sup> Column 1 in Table 4 presents regional inequality measures over time, but where the regional units are defined according to the Bureau of the Census (these include nine geographic units; see footnote to Table 4). The inequality index in column 1 exhibits a definite secular pattern over the whole development spectrum; during the early stages of growth, 1840-80, regional inequality increased or regional divergence was the rule; from 1880 to 1920, the degree of inequality stabilized and even revealed a significant decline; the 1920-60 experience has been varied, to be sure, but generally the evidence suggests a secular decline in the North-South problem, the rate of which has accelerated from the mid-1930's to the present.

It should be noted first that the tendency towards regional divergence, prevalent in our early and mid-19th century history, cannot be explained entirely by the Civil War and the bitter period of reconstruction which followed. Regional divergence was the case *prior* to the Civil War: the tendency towards increasing North-South dualism is evident, although not striking, between 1840 and 1860.<sup>40</sup> The sharpest increase in regional inequality does occur, however, between 1860 and 1880, a period covering both Civil War and Reconstruction.<sup>41</sup>

What can we say about the varied time pattern of American regional inequality between 1900 and 1940? The Great Depression temporarily halted the secular tendency towards a reduction in geographic income differentials because of the relatively severe effects of that period upon the agricultural regions in the United States; by 1940, we had recovered a relatively low degree of regional inequality which had been achieved by 1920. Indeed, Easterlin has shown us that when our data is adjusted by use of National Bureau reference cycle averages, the 1930's do not seriously interrupt the great secular decline in regional dualism which has persisted for almost eighty years.<sup>42</sup> Yet the 1920's still remain a significant exception.

38. It has long been recognized among politicians and economists alike that national depressions or periods of stagnation have inequitable effects upon the distribution of income, and this is true, too, of the regional impact of such periods.

39. See Easterlin in *Trends* (*op. cit.*) for an extensive description of U. S. experience.

40. Nevertheless, it should be pointed out that if the American South is excluded from our regional measurement, regional inequality becomes very slight and, furthermore, the "classic" pattern all but disappears.

41. The estimates of regional income inequality for 1860 are not given here. See Richard A. Easterlin, "Regional Income Trends, 1840-1950," in Seymour Harris, ed., *American Economic History* (New York: McGraw-Hill, 1961), p. 528.

42. Easterlin in *Trends*, *op. cit.*, pp. 93-94.



## REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 179

This unusual decade in American history experienced regional divergence, contrary to the long-run trend towards convergence, and this was hardly a period of stagnation like the decade which followed.

The second column presents our measure of regional inequality where the regional units are the states themselves. This series, based upon smaller regional units, exhibits a higher measure of regional inequality throughout our 19th and 20th century history. The divergence, or increasing North-South dualism, from 1840 to 1880 is clear in both series, however, and the tendency towards increased regional inequality during the 1920's and 1930's is also pronounced in each. In both cases, the decline in regional inequality and tendency towards convergence has been most impressive during the twenty-five years, 1935-60, a period of active federal concern with regional dualism where public transfers to "depressed areas" has been most significant.

This so-called "classic" pattern also seems to hold for Swedish experience. The degree of regional inequality increased sharply during the decade of the 1920's, from 0.440 to 0.539 in 1930.<sup>43</sup> One could argue that this decade is the terminal one for the early and adolescent stage of Swedish growth. Although the period is admittedly short, we might then argue that the increasing regional inequality from 1920 to 1930 is entirely consistent with our hypothesis. This conclusion is strengthened by the high degree of inequality that *was* prevalent during this period of Swedish history (only the current middle-income nations have North-South problems as severe as that which Sweden experienced in the 1920's). It should be emphasized here that the trend towards increasing regional divergence is not the result of the dominating performance of a small sample of regions, but reflects consistent divergence throughout Sweden. With only two major exceptions, all of the counties below the national 1920 average suffered further decline during the decade. Although the Stockholm region was clearly the most dynamic "*pôle de croissance*," all the highly developed Swedish areas show relative improvement, with the exception of Malmöhus län, which underwent a decline in relative income per capita (120.4 to 118.4 percent of the national average).

During the three decades following 1930, and during a period of mature development, Sweden has undergone a tremendous decline in regional dualism:  $V_w$  fell from 0.539 in 1930 to 0.192 in 1961. This aggregate pattern is again strikingly supported by individual performance; over these three decades each of the high income regions underwent a decline relative to the national average, while every Swedish region with a 1930 per capita income less than that of the national average exhibited a trend approaching that average.

Both Italy and Brazil also seem to have undergone experience with regional inequality consistent with our results thus far. Italian income data on a regional basis is available only from 1928. Our index of Italian regional inequality rises from a fairly high level in 1928, 0.313, to an average figure of 0.360 for 1950-60. Throughout the period 1928-60 it was Piemonte, Liguria, and Lombardia which were the leaders in the North, while in the postwar era both Lazio and Emilia-Romagna have joined them.

The decade pattern *within* the 1950's is confused, to be sure, but the relative stability of  $V_w$  between 1951 and 1960 suggests that Italy has reached a plateau with respect to her North-South problem. This would be consistent with her position on the development spectrum. For that matter, it is difficult to determine whether the sharp increase in  $V_w$  from 1938 to the early 1950's is part of a secular trend, peaking in the 1950's, or whether it is due in part to the disproportionate regional effects of World War II. While Piemonte, Liguria, Lombardia,

43. The year 1930 is not one of severe depression, and therefore our estimate of regional inequality should not be excessively biased by the effects of the Great Depression.



Time Series: Ten Nations

| <u>United States</u> |                      |                     | <u>Norway</u>         |       |
|----------------------|----------------------|---------------------|-----------------------|-------|
|                      | $V_w$<br>(by region) | $V_w$<br>(by state) |                       | $V_w$ |
| 1840                 | 0.231                | 0.279               |                       |       |
| 1880                 | 0.321                | 0.355               | 1939                  | 0.424 |
| 1900                 | 0.299                | 0.322               | 1947                  | 0.253 |
| 1919                 |                      | 0.276               | 1952                  | 0.238 |
| 1920                 | 0.291                | 0.331               | 1957                  | 0.233 |
| 1921                 |                      | 0.373               | 1958                  | 0.221 |
| 1929                 |                      | 0.369               | 1959                  | 0.209 |
| 1930                 | 0.338                | 0.389               | 1960                  | 0.186 |
| 1931                 |                      | 0.395               |                       |       |
| 1932                 |                      | 0.410               | <u>United Kingdom</u> |       |
| 1933                 |                      | 0.394               |                       |       |
| 1934                 |                      | 0.369               |                       |       |
| 1935                 | 0.310                | 0.337               |                       | $V_w$ |
| 1936                 |                      | 0.344               |                       |       |
| 1937                 |                      | 0.326               | 1937                  | 0.116 |
| 1938                 |                      | 0.329               | 1949/50               | 0.074 |
| 1939                 |                      | 0.331               | 1954/55               | 0.064 |
| 1940                 | 0.263                | 0.331               | 1959/60               | 0.071 |
| 1941                 |                      | 0.306               |                       |       |
| 1942                 |                      | 0.269               | <u>Netherlands</u>    |       |
| 1943                 |                      | 0.258               |                       |       |
| 1944                 |                      | 0.236               |                       |       |
| 1945                 | 0.211                | 0.227               |                       | $V_w$ |
| 1946                 |                      | 0.236               |                       |       |
| 1947                 |                      | 0.226               | 1938                  | 0.302 |
| 1948                 |                      | 0.214               | 1946                  | 0.151 |
| 1949                 |                      | 0.212               | 1950                  | 0.123 |
| 1950                 | 0.193                | 0.218               | 1955                  | 0.142 |
| 1951                 | 0.194                | 0.213               | 1958                  | 0.128 |
| 1952                 | 0.189                | 0.209               |                       |       |
| 1953                 | 0.191                | 0.212               | <u>Sweden</u>         |       |
| 1954                 | 0.182                | 0.208               |                       |       |
| 1955                 | 0.182                | 0.207               |                       |       |
| 1956                 | 0.184                | 0.211               |                       | $V_w$ |
| 1957                 | 0.184                | 0.208               |                       |       |
| 1958                 | 0.171                | 0.201               | 1920                  | 0.440 |
| 1959                 | 0.172                | 0.196               | 1930                  | 0.539 |
| 1960                 | 0.176                | 0.195               | 1944                  | 0.311 |
| 1961                 | 0.167                | 0.192               | 1950                  | 0.229 |
|                      |                      |                     | 1955                  | 0.180 |
|                      |                      |                     | 1961                  | 0.192 |

## Canas

1926  
1930  
1935  
1940  
1945  
1950  
1955  
1960

1901  
1907  
1911  
1921  
1928  
1931  
1934  
1938  
1950  
1951  
1952  
1953  
1954  
1955  
1956

# REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 181

Table 4 (continued)

| France  |   |  |  |  |  |
|---------|---|--|--|--|--|
|         | $V_w$<br>(Taxable income<br>per family) |  |  | $V_w$<br>(Personal income<br>per capita) |  |
| 1864    |   |  |  | 0 260                                    |  |
| 1938    | 0 658                                   |  |  |  |  |
| 1946    | 0 416                                   |  |  |  |  |
| 1951    |   |  |  | 0 289                                    |  |
| 1954    |   |  |  | 0 249                                    |  |
| 1955/56 |   |  |  | 0 303                                    |  |
| 1958    |   |  |  | 0 299                                    |  |

| Canada |       | Italy |       | Brazil |       |
|--------|-------|-------|-------|--------|-------|
|        | $V_w$ |       | $V_w$ |        | $V_w$ |
| 1926   | 0 176 | 1928  | 0 311 | 1939   | 0 902 |
| 1930   | 0 221 | 1938  | 0 345 | 1947   | 0 693 |
| 1935   | 0 237 | 1951  | 0 363 | 1948   | 0 689 |
| 1940   | 0 220 | 1952  | 0 384 | 1949   | 0 711 |
| 1945   | 0 189 | 1953  | 0 323 | 1950   | 0 712 |
| 1950   | 0 199 | 1954  | 0 331 | 1951   | 0 725 |
| 1955   | 0 192 | 1955  | 0 346 | 1952   | 0 781 |
| 1960   | 0 175 | 1956  | 0 348 | 1953   | 0 701 |
|        |       | 1957  | 0 344 | 1954   | 0 711 |
|        |       | 1958  | 0 340 | 1955   | 0 692 |
|        |       | 1959  | 0 356 | 1956   | 0 690 |
|        |       | 1960  | 0 372 | 1957   | 0 665 |
|        |       |       |       | 1958   | 0 615 |
|        |       |       |       | 1959   | 0 661 |

| Germany |                     |                         |
|---------|---------------------|-------------------------|
|         | $V_w$<br>Old Empire | $V_w$<br>New W. Germany |
| 1900    | 0 220               |                         |
| 1907    | 0 242               |                         |
| 1913    | 0 226               |                         |
| 1926    | 0 181               |                         |
| 1928    | 0 186               |                         |
| 1932    | 0 176               |                         |
| 1934    | 0 164               |                         |
| 1936    | 0 196               |                         |
| 1950    |                     | 0 221                   |
| 1951    |                     | 0 210                   |
| 1952    |                     | 0 211                   |
| 1953    |                     | 0 202                   |
| 1954    |                     | 0 197                   |
| 1955    |                     | 0 196                   |
| 1960    |                     | 0 191                   |

Table 4 (continued)

## Sources:

- (1) United States. The first column, "by region," is computed by using nine regions as defined by Bureau of Census (see footnote to Table 1).  $V_W$  was computed from data in Richard A. Easterlin, "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950," *Trends in the American Economy in the Nineteenth Century*, (Princeton: Princeton University Press, 1960), T-D-1 and T-D-2, 136-37 for the years 1840, 1900, 1920, 1940, and 1950. The years 1935 and 1945 are derived from Charles F. Swartz and Robert E. Graham, "Personal Income by States, 1929-54," *Survey of Current Business*, (September 1960). See footnote to Table 1 for sources of 1951-61 data. The second column, "by states," is from three sources: 1840, 1880, and 1900 are derived from Easterlin, *Trends*, T-A-1, pp. 97-104; 1919-54 are computed by Frank A. Hanna, *State Income Differentials, 1919-1954* (Durham: Duke University Press, 1959), T-3, p. 36; see footnote to Table 1 for sources of the 1955-61 data.
- (2) Netherlands. See footnote to Table 1.
- (3) Norway. See footnote to Table 1.
- (4) Sweden. See footnote to Table 1 for the source of the data used to compute the figures for 1950, 1955, and 1961. For 1920, the income data is from the Statistiska Centralbyrån, *Folkräkningen den 31 December 1920*, IV (Stockholm, 1926), T-F, pp. 38-39, and the population data is from the same source, T-A, pp. 22-23. For 1930, the income data is from the Statistiska Centralbyrån, *Folkräkningen den 31 December 1930*, VIII (Stockholm, 1938), T-Ae, pp. 86-87, and the population data is derived from the same source, T-G, p. 16, and p. xv. See footnote to Table 1 for regional definitions.
- (5) France. See footnote to Table 1 for regional classification and for sources of 1954, 1955/56, and 1958 data. The 1864 data is also from N. Delefortrie and J. Morice, *Les Revenus Départementaux en 1864 et en 1954*, *Recherches sur L'Economie Française*, #1 (Paris, 1959), cols. 30 and 51, pp. 18-19 and 85-86, and is based on personal income per capita. The 1938 and 1946 taxable income per family data is taken from *Etudes et Conjoncture* (September 1949), T-8, pp. 83-84.
- (6) United Kingdom. All these estimates are based on earned income per taxpayer. The 1937 figure is derived by a different regional breakdown than the later years: it includes thirty county groupings and excludes North Ireland. Colin Clark, "The Economic Function of a City," *Econometrica*, XXIII (April 1945), T-5, 104-05. The 1949/50, 1954/55, and 1959/60 estimates are based on earned income per taxpayer figures: for these years the "standard regional classification" is used (including North Ireland) and is from the *95th, 100th, and 105th Reports of the Commissioners of Her Majesty's Inland Revenue* (see footnote to Table 1).
- (7) Canada. See footnote to Table 1.
- (8) Brazil. See footnote to Table 1.
- (9) Italy. See footnote to Table 1 for 1951-60 data. The regional units for 1928 and 1938 are the same as those for the postwar years: Svimez, *Un Secolo di Statistiche Italiane Nord e Sud, 1861-1961*, Capitolo XI (Roma, 1961), T-295, p. 770.
- (10) Germany. See footnote to Table 1 for 1950-60 data and for description of regional units encompassed by West German border. Because of periodic changes in the national boundary, as well as in the nature of the regional units themselves, it would be impractical and not very meaningful to attempt to construct a continuous time series from 1900-60. The "Old" West Germany series includes Schleswig-Holstein, Hannover, Westfalen, Hessen-Nassau, Rheinprovinz, Bayern, Württemberg, Baden, Hessen, Hamburg, Oldenburg, Braunschweig, Bremen, Lippe, and Schaumburg-Lippe (approximately 56 percent of total pre-war Germany in terms of income). The "Old Empire" series includes East Prussia, West Prussia and Posen, Berlin-Brandenburg, Pommern, Schlesien, Provinz Sachsen, Schleswig-Holstein, Hannover, Westfalen, Hessen-Nassau, Rheinprovinz, Sachsen, Württemberg, Baden, and Hamburg. These income and population data (1900-36) are from the following sources: Statistisches Reichsamt, "Das deutsche

Table 4 (con

Volk  
Reich  
Reich  
Statist  
Table  
1875,  
to use  
Journ

Hopetu  
Vanoni Plan,  
postwar era,  
impossible to  
tions regardi  
of growth and

Brazil  
early-mid 19  
vergence has  
not be reliable  
case it seems  
selves quite  
panied by ext  
southern stat  
to the North  
period of dec  
clines in inv  
of Rio Grande  
movement. A  
positions, wh  
declines. It  
North East va  
less encourag  
come per capi

One mig  
years of her  
hand, and mo  
that Brazilian  
1959 of SUDE  
nations. At a  
tended to dev  
yet the five ye  
to the North

44. See espec  
Making 19

Table 4 (continued)

Volkseinkommen vor und nach dem Kriege," *Statistische Nachrichten zur Statistik des Deutschen Reichs*, No. 24 (Berlin, 1932), Table 12, 13, and 16, pp. 72 and 76, Statistisches Reichsamt, *Wirtschaft und Statistik, Neunzehnter Jahrgang* (Berlin, 1936), p. 363, Statistisches Reichsamt, *Statistisches Jahrbuch für das Deutsche Reich* (Berlin, 1932), Table F-16, p. 525. There are also some scattered observations for the years 1854, 1875, and 1890, but the number of German regions was so limited no attempt was made to use them; see S. N. Protopopovich, "The Distribution of National Income," *Economic Journal*, LXXVI (March 1926), 89-92.

Hopefully for such institutions as the *Consiglio Nazionale delle Ricerche* and the spirit of the Vanoni Plan, the Italian North-South differential has at least seemingly stabilised during the postwar era, although the relative importance of public policy as the causative factor is almost impossible to isolate. This evidence on Italian regional dualism suggests optimistic projections regarding the future size of the North-South problem as Italy passes into mature stages of growth and rapidly ascends into high-income classes.

Brazilian experience is less encouraging. Divergence was the case from 1919 to the early-mid 1950's (or perhaps even from the drought of 1877-79), but, surprisingly, slight convergence has been the rule during the short period since. Brazilian regional income data may not be reliable enough to make strong conclusions from such short-term periods, but in any case it seems clear that the conditions which produced these two opposite trends are themselves quite different. From 1919 to 1955 the increase in our aggregate measure  $V_w$  is accompanied by consistent movements in the disaggregates, with very few exceptions. The advanced southern states increased their per capita income relative, while the underdeveloped states to the North suffered relative declines. This consistency of pattern is less true of the recent period of declining  $V_w$ , 1952-60. Four of the five advanced regions did exhibit relative declines in income per capita, but in large part this was due to the extremely rapid development of Rio Grande do Sul. Among the poorer regions there was a considerable lack of conformity of movement. Most of the underdeveloped states showed only slight increases in their relative positions, while Espírito Santo, Mato Grosso, Amazonas, and Para underwent very severe declines. It appears, then, that the North (Central) suffered a considerable decline, while the North East gained only slightly. Thus the mild decline in  $V_w$  over this brief period is all the less encouraging, since it hardly represents a general trend towards convergence in state income per capita levels.

One might suggest that the inverted U<sup>44</sup> traced out by Brazil's passage through the later years of her early development stage may be due entirely to governmental policy, on the one hand, and movements of the external terms of trade, on the other. We might also comment here that Brazilian concern with its North-South problem (explicitly revealed by the inauguration in 1959 of SUDENE by Kubitschek) is highly unusual relative to the historical experience of other nations. At similar levels of national income per capita and development, few countries have tended to devote increased attention to regional dualism at the expense of other national goals, yet the five year plan of SUDENE involves a heavy net transfer of funds from the Center-South to the Northeast as well as external aid from the United States.<sup>45</sup>

44. See especially Hirschman's discussion in *Progress Toward Progress: Studies of Economic Policy Making in Latin America* (The Twentieth Century Fund, 1961) and *ibid.*, pp. 52.

Quantitative information on French historical experience with regional income inequality is very thin, and we shall discuss it only briefly here. We have one observation for 1864 (based on very suspect data), where  $V_w$  is computed as 0.260, while the next comparable observation is not available until *ninety* years later; in 1951, the coefficient of variation is 0.289. We would have predicted that French regional inequality increased significantly during its modern period of development after the mid-19th century.<sup>45</sup> We would also expect a period of convergence to have set in during her mature stage of growth long before the 1950's. The estimates of  $V_w$  for 1864 and 1951 are at least consistent with the notion.

If we utilize evidence which is less direct and which does not involve regional income data, we find strong support for the hypothesis that France has experienced the "classic" pattern of regional income distribution over the past one hundred years. In an excellent study of internal migration patterns, L. M. Goreux presents data which is extremely pertinent to our problem. Goreux computed coefficients of variation based upon the regional wages of agricultural laborers and also upon regional indices of agricultural production per male laborer. Although his measures of regional inequality are less universal than ours, since they are based only on the agricultural sector, they may still prove to be helpful.<sup>46</sup>

France  
( $V_w$  based on agricultural wages, including  
room and board, by department)

|      |       |      |       |
|------|-------|------|-------|
| 1862 | 0.285 | 1948 | 0.109 |
| 1882 | 0.303 | 1949 | 0.100 |
| 1892 | 0.368 | 1950 | 0.158 |
| 1929 | 0.225 | 1951 | 0.140 |
| 1938 | 0.201 | 1952 | 0.160 |
| 1947 | 0.095 | 1953 | 0.160 |

( $V_w$  based on agricultural product per male  
agricultural laborer by department)

|      |       |
|------|-------|
| 1882 | 0.427 |
| 1910 | 0.406 |
| 1929 | 0.391 |
| 1948 | 0.342 |

45. The lone observation for 1864 does suggest a contradiction of the popular view among European economic historians that France was typified by "polarization" and regional inequity long before her industrial revolution, since the computed  $V_w$  of 0.260 for 1864 does not indicate severe dualism. Regional concentration may still, of course, have been the case.

46. These figures are from Goreux, *op. cit.*, pp. 331 and 343, respectively. The observations are weighted by the distribution of the agricultural labor force by department. Given the information which follows in section IX, a true estimate of  $V_w$  based upon regional *income* inequality would be somewhat lower in level but similar in movement to those estimated by Goreux based only upon conditions in the agricultural sector.

If this data can be interpreted without qualification, then it suggests that regional inequality became increasingly severe from the 1860's to World War I. Thereafter, secular convergence appears to have been the rule, excluding the temporary interruption of World War II and reconstruction. This pattern is not only consistent with the experience of other presently developed nations, but it is extremely similar with the time path of German regional distribution (see below).

Canada does not reveal any significant trend towards either divergence or convergence during the thirty-five year period, 1926-60, for which regional income data are available. Given Canada's size and relative immaturity compared with the United States,  $V_w$  was surprisingly low in the 1920's, and the index of regional inequality for 1960 was almost precisely that of 1926. Recognizing that the increasing level of  $V_w$  to the 1930's simply reflects the effects of the depression on the Prairie provinces and that the decline in  $V_w$  thereafter reflects a reversal of those conditions, the Canadian case remains a curious one still. If we artificially separate the period into 1926-35 and 1935-60, the disparate patterns become striking. In the earlier period, the "backward" eastern provinces show a mixed performance, with New Brunswick and Prince Edward Island just barely maintaining their relative income per capita positions, while Nova Scotia and Quebec reveal an impressive improvement. The severe effects of the depression upon the Prairie provinces, however, dominates our aggregate measure (e.g., Saskatchewan declines from 102.4 to 63.1 percent of the national average), and increasing inequality is the general rule. The same lack of consistency among the underdeveloped regions appears in the latter period as well. The Prairie provinces undergo impressive improvement, while the Maritime provinces just barely hold their own and Quebec suffers a significant decline. In summary, apart from the interlude of the 1930's, stability in  $V_w$  has been the rule, with Quebec and the Maritime provinces barely maintaining growth rates equivalent to those of Ontario and British Columbia, while the Prairie provinces reveal extreme instability producing fluctuations around the national average. We will say more about Canadian experience and its dissimilarity with U.S. history in the next section.

The German data does not cover the largest portion of her earlier stages of modern development, but begins only with 1900. Generally,  $V_w$  did experience a minor increase during the decade and a half prior to World War I. Our hypothesis would suggest that this period would be a terminal one following four decades of development and concomitant regional divergence. This seems extremely unlikely, given the low level of regional inequality which existed in 1900 (less than or equal to the  $V_w$ 's of those nations currently in Kuznets' income class II). Furthermore, the mild increase in  $V_w$  prior to World War I does not describe a period in which all the developed states are growing at rates exceeding the national average. Only Berlin-Brandenburg shows a significant improvement in relative income per capita, while Hamburg, Hessen-Nassau, Rhein province, and Saxony all reveal stability. Nevertheless, regional inequality declined fairly consistently from 1907 to the mid-1930's. The puzzling phenomena is not so much the *time pattern* of German regional inequality, but its *low level* throughout the period 1900-60. None of the other European countries appears to have had similar experience with geographic income differentials. It is interesting to note that German experience with *size* distribution of income is strikingly similar to what we have already described as her experience with regional inequality. Kuznets' estimates of German size distribution reveal fairly low indices of concentration during the late 19th and early 20th centuries compared with other nations at similar stages of development. Furthermore, it appears that Germany underwent increasing inequality of size distribution up to the period 1896-1912.<sup>47</sup>

47. Kuznets, "Quantitative Aspects of the Economic Growth of Nations. VIII. Distribution of Income by Size," *op. cit.*, Table 16, pp. 60-62.

With the exception of Germany and perhaps Canada, what slim historical evidence we do have seems to be at least consistent with the results of our cross-section analysis.<sup>48</sup>

It would be of major interest to us to know more about the experience of planned economies with regional inequality, although any comparative study involving the East European or Soviet economies would involve questionable indirect evidence and conjecture. It seems highly unlikely that the Communist nations have sacrificed rapid national growth for the "secondary" Marxian goals of (1) introducing industrialization throughout the country in order to achieve the necessary conditions for socialism on a nationwide scale and (2) achieving idealistic equalitarianism implied by the socialist society.<sup>49</sup> In the case of Poland, what little evidence we have concerning regional resource-allocation suggests that goals of reducing regional dualism have been subordinated to national development goals, and that increasing regional divergence has been the case in the postwar period of early industrialization. It appears that Poland has been concentrating her incremental resources in the Upper Silesian Industrial District.<sup>50</sup> Furthermore, now that Soviet Russia has reached a relatively mature stage of growth (equivalent to the middle-income nations or higher), has that nation undergone any tendency towards convergence in regional development levels and reduction in regional dualism? It may turn out that Russia, given its size and income level prior to World War II, actually did not undergo as sharp a movement in regional divergence as, say, Brazil. This seems a likely supposition, given Russian difficulties with inland transportation during the 1930's and increasing stress upon regional self-sufficiency, as well as the military insistence upon regional decentralization. Whatever the case, we do know that the Seven Year Plan in 1959 included in it significant regional goals, and also the 1956 movement toward decentralization itself may imply a serious attempt to reduce regional inequities generated by the fabulous growth of three decades. Finally, we have already seen that in spite of official pronouncements and alleged effort, Yugoslavia underwent increasing regional divergence between 1956 and 1960 (see Appendix and Table 2).

#### VI. The Historical Patterns: Two

We did attempt to extend our historical evidence by the substitution of a variable which would approximate regional income per capita or level of development. By using this very imperfect substitute, the share of agricultural laborers in the total labor force by regions, we were able to lengthen considerably our time series for Canada, Italy, and Brazil. We derived

48. Quantitative evidence describing this aspect of British economic history is not available. However, Arthur Redford in his classic study on British 19th century interregional migration suggests that the development of Great Britain's industrial centers in the first part of that century tended to increase regional dispersion in wage rates. Arthur Redford, *Labour Migration in England, 1800-50* (Manchester: Manchester University Press, 1926), and see Goreux's comments in *op. cit.*, p. 343. Commenting on Spanish experience with regional development, Lasuen implies that regional divergence was initiated with the early development of the heavy metal industries in the Bilbao area and the textile industries in the Barcelona area, both of which began their regional development almost fifty years ago. Furthermore, secular divergence may still be the case, for "although a little spreading has taken place (mainly around Barcelona, less so around Bilbao) the backwash effects are probably stronger than ever." With regard to the last phrase, we have already expressed some disagreement (see section III). Lasuen, *op. cit.*, p. 177.

49. Dziewonski, *op. cit.*, p. 45.

50. *Ibid.*, pp. 43-57.



a rough index of regional inequality by using the square of the differences between regional shares of agricultural employment in the labor force (hereafter termed  $A/L$ ) and that of the nation as a whole.<sup>51</sup>

A glance at Table 5a gives an idea of just how imperfect, as an approximation of income per capita,  $A/L$  is in computing inequality indices. In this limited cross-section sample, the rank correlation (Spearman's coefficient) between our  $V_w$  based on income per capita, and  $\Delta_w$  based on  $A/L$ , is hardly very impressive:  $r = 0.576$ , and when the major offender, Brazil, is eliminated,  $r = 0.758$ . Although  $A/L$  may be a poor substitute as a measure of regional income inequality, this should not imply at the same time that the correlation between  $A/L$  and income per capita is poor between regions and *within* nations. On the contrary, for all nations which have such data available, regional income per capita and the  $A/L$  share revealed highly significant inverse correlations similar to the results of the Chenery-Kuznets-Clark *international* cross-sectional studies.<sup>52</sup> Nevertheless, the slope of the function estimating the relationship between regional  $A/L$  and income per capita varies considerably between countries. In the cases of Finland and Austria, the regional variations in  $A/L$  are much more wide than those of income per capita compared with such nations as Brazil, Italy, and Sweden. To put it in another way, for some nations (most notably Brazil) regional disparities in agricultural productivity are almost as important as the regional role of manufacturing employment in explaining geographic differences in income per capita levels.<sup>53</sup>

51. More precisely, this index of regional inequality,  $\Delta_w$ , is the following:

$$\Delta_w = \sqrt{\sum_i [(A/L)_i - (A/L)]^2 \cdot \frac{f_i}{n}} \times 100$$

where  $(A/L)_i$  = share of agricultural labor in total labor force of  $i^{\text{th}}$  region,  
 $(A/L)$  = share for the nation as a whole,  
 $f_i$  = population of the  $i^{\text{th}}$  region,  
 $n$  = national population.

52. Assuming a simple linear relationship of the form

$$\frac{A}{L} = \beta_0 + \beta_1 Y$$

where  $\frac{A}{L}$  is the proportion of the labor force employed in agriculture in each region, and  $Y$  is the income per capita of each region, we get the following results:

| Country       | No. Regions | $\hat{\beta}_1$  |
|---------------|-------------|------------------|
| Great Britain | 10          | -0.1587 (0.0495) |
| Austria       | 9           | -0.7071 (0.0546) |
| Sweden        | 25          | -0.0266 (0.0029) |
| Brazil        | 20          | -0.2126 (0.0675) |
| Italy         | 16          | -0.4024 (0.0514) |
| Canada        | 9           | -0.3272 (0.0571) |
| Finland       | 23          | -0.8514 (0.0569) |

53. The  $\hat{\beta}$  coefficients above do not precisely show this. It would appear that Sweden and Great Britain exhibit even less variation in  $A/L$  relative to income per capita variation than Brazil. Clearly, the absolute importance of the agricultural sector is small for all regions in Sweden and Great Britain, while this is not the case for Brazil. This interesting topic is pursued further in Section IX.



TABLE 5a.

Index of Regional Inequality Using Agricultural Labor Force  
As a Share in Total Labor Force: Cross-Section

| Country       | Year | $V_w$<br>(using $Y_{pc}$ ) | $\Delta_w$<br>(using $A/L$ ) |
|---------------|------|----------------------------|------------------------------|
| Italy         | 1951 | 0.363                      | 14.22                        |
| Brazil        | 1950 | 0.732                      | 13.10                        |
| Canada        | 1951 | 0.192                      | 10.19                        |
| Finland       | 1950 | 0.331                      | 23.20                        |
| Great Britain | 1951 | 0.141                      | 5.31                         |
| Austria       | 1957 | 0.225                      | 15.55                        |
| Spain         | 1957 | 0.387                      | 22.69                        |
| United States | 1950 | 0.218                      | 9.46                         |
| Japan         | 1959 | 0.259                      | 16.59                        |
| Sweden        | 1940 |                            | 14.12                        |
|               | 1944 | 0.311                      |                              |

TABLE 5b.

Time Series For Canada, Italy, and Brazil:  $\Delta_w$

| Canada |       | Italy |       | Brazil |       |
|--------|-------|-------|-------|--------|-------|
| 1901   | 7.14  | 1861  | 6.55  | 1920   | 7.76  |
| 1911   | 9.88  | 1871  | 7.88  | 1940   | 8.20  |
| 1921   | 12.35 | 1881  | 6.76  | 1950   | 13.10 |
| 1931   | 12.68 | 1901  | 7.84  |        |       |
| 1941   | 12.60 | 1911  | 9.41  |        |       |
| 1951   | 10.19 | 1921  | 10.94 |        |       |
|        |       | 1931  | 12.14 |        |       |
|        |       | 1936  | 12.72 |        |       |
|        |       | 1951  | 14.22 |        |       |

Source of labor force data:

- (1) Brazil. *The Development of Brazil*, Joint Brazil-United States Economic Development Committee (Washington, D. C., 1953), Tables VIII and XI, pp. 291-92.
- (2) Italy. Svimez, *Cento Anni Di Statistiche Sulle Regioni D'Italia* (Rome, 1961), Table 10, pp. 18-22.
- (3) Canada. Dominion Bureau of Statistics, *Census of Canada, 1951* (Ottawa, 1953), Vol. IV, Table 2, and Vol. 1, Table 1.

Table 5b (continued)

Table 5b (continued)

- (4) Spain. Banco de Bilbao, *Renta Nacional de España y su Distribucion Provincial*, 1957 (Bilbao, 1958), pp. 46-47.
- (5) United States. *U. S. Census of Population, 1950* Vol. II, Part I, Table 83.
- (6) Japan. See footnote to Table 1.
- (7) Austria. "Die Verteilung des Volkseinkommens nach Bundesländern," *Monatsberichte des Oesterreichischen Institutes für Wirtschaftsforschung*, Supplement No. 60 (December 1959), Table 15, p. 17.
- (8) Finland. Lars Wahlbeck, *Om Inkomstnivåns Geografi i Finland år 1950*, II, Ekonomi och Samhälle, Skrifter utgivna av Svenska Handelsskolekolan, No. 2 (Helsingfors: Söderström and Co., 1955), Table 11, pp. 576-77.
- (9) Sweden. Statistiska Centralbyrån, *Statistisk Årsbok för Sverige, 1945* (Stockholm, 1945), Table 27, pp. 36-37.
- (10) Great Britain. *Census of England and Wales, 1941*, Occupational Tables (London: Her Majesty's Stationary Office, 1956), Table 20, pp. 152-67.

In spite of these qualifications, we did make use of the A/L index,  $\Delta_w$ , to extend our quantifiable historical series backwards for the three nations exhibited in Table 5b. It is interesting to note that Brazil *had* been undergoing divergence in regional income levels for two decades prior to 1940, while the most violent increase in regional dualism occurred during her modern era of industrialization, 1940-50 (the data for the decade 1950-60 was not available to us). This again appears to support our hypothesis. The use of A/L data also helps solve some of the mystery surrounding Canadian historical experience:  $\Delta_w$  increased rapidly during a very impressive period of Canadian growth, from 1901 to World War I (or more accurately, to 1921).<sup>54</sup> Stability in regional inequality was indeed the case from 1921 to 1941, and the decline since World War II does appear to be part of a secular trend, rather than a short-term movement. Finally, the movement of  $\Delta_w$  in Italy from unification to the modern postwar era seems consistent with the "classic" relationship between regional inequality and national development. There is only a mild increase in regional dualism from unification to the turn of the century:  $\Delta_w$  increases from a low level in 1861, 6.55, to 7.84 by 1901. The rapid increase in regional inequality occurs during the first really impressive period of modern Italian growth from the late 19th century to World War I. Incidentally, if we heroically assume that North-South differentials in labor productivity were the same in Italy in 1861 as in the United States in 1950, it would appear that the North-South problem was less serious in Italy at the turn of the century. Keeping in mind the restrictiveness of our assumptions concerning productivity, it would seem that there is a great deal of truth to the contention that serious Italian regionalism was not inherited at the time of unification, nor was it significantly increased by governmental policy during the last four decades of the 19th century.<sup>55</sup> Whatever the case,  $\Delta_w$  increased continually after 1921, but at a slower rate. All of this appears to be consistent with our qualitative evidence concerning Italian regional development during the national growth process.

#### VII. Measurement: Economic Significance or Political Reality?

There are a number of alternative statistical measures one can use for determining the extent of regional inequality and its change over time. The preference for an unweighted index over a weighted one, we think, is indefensible. The choice of an index which squares

54. It might prove fruitful to examine the nature of the tremendous inflow of foreign capital and labor into Canada during the period to learn more about the impact of those factor imports upon regional inequality.

55. See, for instance, Eckaus, *op. cit.*

regional deviations about the national mean is less clear. In an earlier section we defended our use of the former, since it was helpful in analysis of variance which we pursued elsewhere in the research; in any case, the behavior of  $M_w$  and  $V_w$  is so similar in both cross-section and time series that the discussion becomes academic. More serious is the choice of our measure of regional levels of development in computing aggregate indices of regional inequality. Although an index which is based upon regional income per capita relatives may have more meaning in understanding the process of regional inequality over the development spectrum, it may not be *politically* meaningful. It is quite possible and hardly uncommon that a period of convergence in regional income per capita *relative to a national average* may at the same time be one of increasing absolute differentials. An index based on the former will be determined by regional growth differentials; one based on the latter will be influenced by a mixture of regional growth rate differentials and initial absolute differentials.

If political decision-makers are indeed motivated by absolute differentials in regional income, then a comparison of our index computed from income relatives with that computed from absolute differentials might be helpful. The empirical evidence below compares the movement in  $M_w$ , a weighted mean deviation based upon income relatives, and in  $M_w^a$ , a weighted mean deviation based upon absolute income differentials.<sup>56</sup> To illustrate the divergent movements we have used the recent experience of the United States, Canada, Sweden, and Italy as examples.

Comparative Behavior of  $M_w$  and  $M_w^a$

| <u>United States</u> |                         |                           | <u>Canada</u> |                         |                           |
|----------------------|-------------------------|---------------------------|---------------|-------------------------|---------------------------|
|                      | <u><math>M_w</math></u> | <u><math>M_w^a</math></u> |               | <u><math>M_w</math></u> | <u><math>M_w^a</math></u> |
| 1951                 | 17.6                    | 263.7                     | 1950          | 17.8                    | 169.3                     |
| 1955                 | 17.0                    | 281.2                     | 1955          | 17.2                    | 175.8                     |
| 1961                 | 14.8                    | 264.4                     | 1960          | 15.5                    | 169.2                     |

| <u>Italy</u> |                         |                           | <u>Sweden</u> |                         |                           |
|--------------|-------------------------|---------------------------|---------------|-------------------------|---------------------------|
|              | <u><math>M_w</math></u> | <u><math>M_w^a</math></u> |               | <u><math>M_w</math></u> | <u><math>M_w^a</math></u> |
| 1951         | 30.3                    | 519.4                     | 1950          | 17.7                    | 617.3                     |
| 1955         | 30.1                    | 610.2                     | 1955          | 14.0                    | 576.8                     |
| 1960         | 32.4                    | 846.7                     | 1961          | 14.9                    | 671.3                     |

Source: See Tables 1 and 2.

56. To be more precise,

$$M_w^a = \frac{\sum \frac{|y_i - \bar{y}| f_i}{n}}{P}$$

where  $y_i$  = income per capita of the  $i^{\text{th}}$  region,  
 $\bar{y}$  = national income per capita,  
 $n$  = national population,  
 $f_i$  = population of the  $i^{\text{th}}$  region,  
 $P$  = index of the general price level.

It should be emphasized that  $M_w^a$  cannot be used for between-country comparisons, since money incomes are deflated only by a general national price index—they have not been converted into common currency units.

As the reader can verify by looking at the accompanying table, although the United States and Canada have both recently undergone considerable convergence in regional income per capita relatives, neither have had any success in reducing absolute differentials between regions. The poorer regions have managed to grow at rates so much higher than the richer ones, however, that absolute differentials have stabilized. Given that in 1951, for instance, the Northeastern region of the United States had a per capita income level half again as large as that of the South, the stability in the absolute differential represents quite a considerable effort. The same description appears to hold for Sweden as well. Given much higher initial regional differentials for Italy, the divergence between  $M_w$  and  $M_w^S$  is much sharper. Although Italian regional inequality based on income relatives declined slightly between 1951 and 1955, absolute income per capita differentials increased sharply.

To summarize this brief section, we have tried to show that, to have any economic meaning and to be useful in explaining the behavior of this aspect of the growth process, an analysis of regional inequality and geographic dualism must be based upon income relatives and thus upon growth rate differentials. We recognize that absolute income differentials may have more political meaning, but to expect that the regional convergence typical of national maturity will also produce reductions in these absolute differentials is to expect a great deal indeed. Thus it would be folly to assume that the strain of economic development upon regional dualism will lessen as the young nation moves into self-sustained growth or into income class IV, for even the highly integrated American economy, with its trend towards regional convergence, has not been able to reduce the *absolute* gap between North and South.

#### VIII. Population Redistribution versus Income Growth

Recall that our index of regional inequality has two components. First, the index is an aggregate measure of the dispersion of regional levels of per capita income (or  $A/L$ ) about the national mean. Second, each regional observation is weighted by its importance, that is, by its share in the national population. The question therefore arises of the relative contribution of changes in regional per capita income over time versus the contribution of population redistribution in producing these historical variations in  $V_w$ . Is it differentials in regional per capita income growth which generally dominate our measure of regional inequality, or does population redistribution and changing regional population weights play a significant role?

It should be made clear immediately what we do *not* intend to investigate here. It should be obvious that internal migration may have significant effects on the regional distribution of the national population over time. But internal migration also should effect wages and income per capita in both the sending and receiving region. In this section, we are implicitly assuming that population redistribution has *no* causative effect upon income per capita in the regions themselves. Our goal is therefore a much more limited one. We are asking whether changes in regional population weights over time (due either to differential natural rates of growth, internal migration, or external migration) significantly effect the historical pattern of  $V_w$  in the course of national development.

To measure the relative importance of population redistribution versus income growth we have used analysis of variance.<sup>57</sup> Changes in the weighted variance of regional income per capita about the national mean can be artificially decomposed into three separate components: changes in variance due to shifting population weights, changes in variance due to divergent regional income per capita growth, and, an indefinable component necessary to preserve

57. We would like to thank Alfred Conrad for the suggestions he made pertaining to this measurement problem.

additivity, changes in variance due to the interaction of both income and population change.<sup>58</sup> The results of these computations are given in Table 6. Ten countries were examined for different time periods in their growth experience, and in three cases the A/L data was used in addition to the income data.

In Table 6 each of these three components is given as a percentage of the change in total variance. Using Italy as an example, and using the available income data, during the period 1951-60, 29.0 percent of the change in total regional variation appears to be due to population redistribution or changes in regional population weights. This turns out to be an unusual case, since the historical experience of most nations has been that population redistribution has a relatively insignificant effect upon changes in regional inequality of income distribution. Indeed, in the twenty-one cases exhibited in Table 6, nine show that changes in variance due to population redistribution acted in a fashion opposite to that of changes in total variance.<sup>59</sup> For that matter, only seven of the twenty-one cases exhibit population redistribution playing a significant role; but in only one case, India, does population redistribution dominate changes in total variation. In all other cases disparity in regional per capita income growth is the major explanation of variations in  $V_w$ , regional income inequality.

To repeat, no inference should be drawn concerning the effect of internal labor migration upon regional inequality, since we would expect labor migration to effect income per capita levels as well as to change regional population weights.

58. Between two time periods  $t = 0$  and  $t = 1$ , the increase or decrease in total regional variance can be decomposed in the following fashion:

$$\sum_i (y_i^1 - \bar{y}^1)^2 f_i^1 - \sum_i (y_i^0 - \bar{y}^0)^2 f_i^0 = \sum_i (y_i^0 - \bar{y}^0)^2 (f_i^1 - f_i^0) + \sum_i f_i^0 [(y_i^1 - \bar{y}^1)^2 - (y_i^0 - \bar{y}^0)^2] \\ + \sum_i (f_i^1 - f_i^0) [(y_i^1 - \bar{y}^1)^2 - (y_i^0 - \bar{y}^0)^2]$$

where  $y_i^1$  is the income per capita of the  $i^{\text{th}}$  region in  $t = 1$

$\bar{y}^0$  is the income per capita of the nation in  $t = 0$

$f_i^1$  is the share of the  $i^{\text{th}}$  region's population in the population for  $t = 1$

Obviously what we have done was to fix regional income differentials at levels existing in the initial period and then allowed the population weights to vary; similarly, we fixed population weights at those of the initial period and then allowed variations in regional income growth to occur; finally, both were then allowed to vary over the time period. These three components should then sum up to total change in regional variation between  $t = 0$  and  $t = 1$  computed independently.

59. For example, Italian experience between 1901 and 1936 was such that population redistribution tended to *diminish* regional inequality while the measure of total regional inequality *increased*. The eight other cases are Canada (1926-33), Brazil (1952-59 and 1920-50), France (1864-1954), Germany (1907-36), Sweden (1944-61), Norway (1939-60), and the United States (1880-1920). There is no consistency, furthermore, with regard to where these time periods fall on the inverted "U": for Brazil it is during a period of rising regional inequality, for Sweden a period of declining inequality, and for the United States a period of relative stability.

## REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 193

TABLE 6.

Decomposition of Variance: Population Versus Income  
Growth for Ten National Time Series

| Country          | Absolute<br>variance<br>(1) | %<br>variance <sup>e</sup><br>(2) | Absolute<br>variance<br>(3) | %<br>variance<br>(4) | Absolute<br>variance<br>(5) | %<br>variance<br>(6) |
|------------------|-----------------------------|-----------------------------------|-----------------------------|----------------------|-----------------------------|----------------------|
| <u>1951-60</u>   |                             |                                   |                             |                      |                             |                      |
| Italy            | 18.7992 <sup>a</sup>        | 29.0                              |                             |                      |                             |                      |
| (Income)         | 44.7204 <sup>b</sup>        | 68.9                              |                             |                      |                             |                      |
|                  | 1.4048 <sup>c</sup>         | 2.2                               |                             |                      |                             |                      |
|                  | 64.9244 <sup>d</sup>        |                                   |                             |                      |                             |                      |
| <u>1861-1901</u> |                             |                                   |                             |                      |                             |                      |
| Italy            | 1.9874                      | 3.7                               | <u>1901-36</u>              |                      |                             |                      |
| (A/L)            | 58.7439                     | 109.0                             | - 2.5432                    | - 0.5                |                             |                      |
|                  | - 6.8552                    | - 12.7                            | 517.3368                    | 100.3                |                             |                      |
|                  |                             |                                   | 1.2250                      | 0.2                  |                             |                      |
|                  | 53.8761                     |                                   | 516.0186                    |                      |                             |                      |
| <u>1926-33</u>   |                             |                                   |                             |                      |                             |                      |
| Canada           | - 7.567                     | - 1.8                             | <u>1933-48</u>              |                      |                             |                      |
| (Income)         | 427.991                     | 100.6                             | - 48.650                    | 10.7                 | <u>1951-61</u>              |                      |
|                  | 5.180                       | 1.2                               | -462.201                    | 101.4                | -10.396                     | 11.5                 |
|                  |                             |                                   | 54.895                      | - 12.0               | -82.146                     | 91.2                 |
|                  | 425.604                     |                                   | -455.956                    |                      | 2.463                       | - 2.7                |
|                  |                             |                                   |                             |                      | -90.079                     |                      |
| <u>1911-31</u>   |                             |                                   |                             |                      |                             |                      |
| Canada           | 87.34                       | 9.9                               |                             |                      |                             |                      |
| (A/L)            | 695.69                      | 78.8                              |                             |                      |                             |                      |
|                  | 99.92                       | 11.3                              |                             |                      |                             |                      |
|                  | 882.95                      |                                   |                             |                      |                             |                      |
| <u>1939-52</u>   |                             |                                   |                             |                      |                             |                      |
| Brazil           | 98.440                      | 2.8                               | <u>1952-59</u>              |                      |                             |                      |
| (Income)         | 3,311.903                   | 92.6                              | 140.111                     | - 8.2                |                             |                      |
|                  | 166.642                     | 4.7                               | -1,790.994                  | 105.1                |                             |                      |
|                  | 3,576.985                   |                                   | - 51.792                    | 3.0                  |                             |                      |
|                  |                             |                                   | -1,702.675                  |                      |                             |                      |
| <u>1920-50</u>   |                             |                                   |                             |                      |                             |                      |
| Brazil           | - 1.6390                    | - 2.1                             |                             |                      |                             |                      |
| (A/L)            | 77.9969                     | 99.5                              |                             |                      |                             |                      |
|                  | 2.0165                      | 2.6                               |                             |                      |                             |                      |
|                  | 78.3744                     |                                   |                             |                      |                             |                      |
| <u>1864-1954</u> |                             |                                   |                             |                      |                             |                      |
| France           | 477.9323                    | -613.7                            |                             |                      |                             |                      |
| (Income)         | -273.7338                   | 351.5                             |                             |                      |                             |                      |
|                  | -282.0777                   | 362.2                             |                             |                      |                             |                      |
|                  | - 77.8792                   |                                   |                             |                      |                             |                      |

Table 6 (continued)

|                      |             |                  |           |                |
|----------------------|-------------|------------------|-----------|----------------|
| <u>1949/50-59/60</u> |             |                  |           |                |
| Australia            | 0.1064      | 0.7              |           |                |
| (Income)             | 14.4680     | 99.9             |           |                |
|                      | - 0.0889    | - 0.6            |           |                |
|                      | 14.4855     |                  |           |                |
| <br>                 |             |                  |           |                |
| <u>1907-36</u>       |             | <u>1950-60</u>   |           |                |
| Germany              | 6.8119      | - 3.3            | - 1.4930  | 1.2            |
| (Income)             | -196.3604   | 96.4             | -122.3828 | 100.9          |
|                      | - 14.0842   | 6.9              | 2.6185    | - 2.2          |
|                      | -203.6327   |                  | -121.2573 |                |
| <br>                 |             |                  |           |                |
| <u>1944-61</u>       |             |                  |           |                |
| Sweden               | 52.4709     | - 8.7            |           |                |
| (Income)             | -614.6234   | 103.0            |           |                |
|                      | - 34.6402   | 5.8              |           |                |
|                      | -596.7927   |                  |           |                |
| <br>                 |             |                  |           |                |
| <u>1939-60</u>       |             |                  |           |                |
| Norway               | 68.3723     | - 8.1            |           |                |
| (Income)             | -931.9665   | 109.8            |           |                |
|                      | 14.8386     | - 1.7            |           |                |
|                      | -848.7556   |                  |           |                |
| <br>                 |             |                  |           |                |
| <u>1930-61</u>       |             |                  |           |                |
| United -             | 47.0053     | 4.2              |           |                |
| States -1,           | 107.2758    | 98.8             |           |                |
| (Income              | 34.1195     | - 3.0            |           |                |
| by state)            | -1,120.1616 |                  |           |                |
| <br>                 |             |                  |           |                |
| <u>1840-60</u>       |             | <u>1880-1920</u> |           | <u>1920-30</u> |
| United               | 12.972      | 12.1             | 315.58    | - 44.0         |
| States               | 175.930     | 164.7            | -668.24   | 93.2           |
| (Income              | - 82.067    | - 76.8           | -364.45   | 50.8           |
| by region)           | 106.835     |                  | -717.11   |                |
|                      |             |                  |           | 22.81          |
|                      |             |                  |           | 277.41         |
|                      |             |                  |           | - 8.05         |
|                      |             |                  |           | - 2.8          |
|                      |             |                  |           | 292.17         |
| <br>                 |             |                  |           |                |
| <u>1950/51-60/61</u> |             |                  |           |                |
| India                | 465.41      | 50.8             |           |                |
| (Income)             | 204.92      | 22.4             |           |                |
|                      | 245.81      | 26.8             |           |                |
|                      | 916.14      |                  |           |                |

- Variance attributable to population redistribution.
- Variance attributable to income per capita (or A/L) changes.
- Variance attributable to both population and income changes. See text.
- Total variance.
- This column gives a, b, and c as a percentage of d.

Table 6 (continued)

Note: The data used in deriving the variance estimates above are taken from the same sources as indicated in earlier tables. The reader will note further that in three cases A/L has been used as the "income" variable. Finally, the regional units underlying the estimates in this table are those used in Table 1.

#### IX. Regional Labor Participation and Sectoral Distribution

This section involves two additional steps which attempt to increase further our understanding of the relationship between regional inequality and economic development. The first attempts to isolate the role of regional variations in labor participation rates and their contribution to spatial differences in income per capita levels. Given significant geographic income per capita differentials, one would anticipate high rates of participation in the rich North and low rates in the South for much the same reasons that labor participation rates tend to be relatively low in low-income countries.<sup>60</sup> The question then arises, how much of these observed regional inequalities are explained by productivity differentials and how much by participation rates?

Column 7 in Table 7 throws some light on this question. Here we have a small sample of thirteen countries. Column 3 exhibits the computed index of inequality using regional income per capita and weighting by regional population shares. Column 4 presents a different index of inequality based on labor productivity: this index measures regional variation in income (or product) per worker, and each regional variation is weighted by regional labor force shares in the national labor force.<sup>61</sup> Column 7 is simply a ratio of the inequality index based on income per capita to that which is based on labor productivity. At a variety of national development levels and in all cases but two, Japan and the United States (1900), the inequality index is lower when computed from regional productivity data. Labor participation rates appear to play a significant role in explaining regional dualism at all levels of national development.

We might interject the remark that the positive correlation between income levels and labor participation rates tends to be stronger between regions within national boundaries than between nations themselves (two exceptions are discussed below). Furthermore, the range of variation in regional labor participation rates is apparently greater than between nations. Using Kuznets' post-war data,<sup>62</sup> the range in labor participation rates (including

60. The correlation between national levels of development and labor participation rates is, however, far from perfect.

61. As in Section III,

$$V_w = \frac{\sqrt{\sum_i (y_i - \bar{y})^2 \frac{f_i}{n}}}{\bar{y}}$$

but where  $f_i$  = labor force of  $i^{\text{th}}$  region,  
 $n$  = national labor force,  
 $y_i$  = income per worker in the  $i^{\text{th}}$  region,  
 $\bar{y}$  = national income per worker.

The indices in columns 3 and 4 use the same regional income or product data.

62. Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations. II. Industrial Distribution of National Product and Labor Force," *Economic Development and Cultural Change*, V, No. 4 (July 1957), Appendix Table 8, pp. 106-07.



unpaid family labor and expressed in percentages) is something like the following: France—51.5; United Kingdom—46.2; Germany—46.3; and the United States—39.8; compared with Mexico—32.4; Chile—36.9; Brazil—33.0; and Egypt—37.6. Contrast this with the range of regional labor participation rates in, for example, Italy and Sweden: in Italy (1951) the range lies between 48.0 (Piedmont) and 50.0 (Venice), on the one hand, and 33.0 (Sicily) and 34.0 (Sardinia), on the other; in Sweden (1944) the range lies between 55.1 (the Stockholm region) and 48.6 (Malmöhus), on the one hand, and 39.9 (Norrbottens) and 42.2 (Västerbottens), on the other. The suggestion here, of course, is that in the light of the much smaller range in regional development levels, the range in regional participation rates appears to be significantly larger. Does this suggest that the higher rate of internal, relative to external, migration plays a consistent role in tending to generate regional labor participation rate differentials due to its selective nature?<sup>63</sup>

Note, too, the effect that changes in regional participation rates can have upon historical experience of national economies with spatial income per capita differentials. Although Italian interregional disparities in income per capita increased slightly between two isolated postwar years, 1951 and 1960, productivity disparities declined! If participation rates had remained unchanged during the 1950's, Italian attempts to reduce the North-South schism would have shown more notable success. Notice also that the perverse change in regional participation rates in Canada has dampened what might otherwise have been a very sharp decline in Canadian regional income disparities during the two decades 1931-51. The opposite appears to have been the case for Sweden from 1944 to 1960.

In summary, not all of the North-South problem in these countries is due to productivity differentials: systematic regional differentials in age structure patterns and the like tend to further widen the regional gap already produced by productivity differentials. As we have already noted, Japan is one exception to the rule (as she is to almost any economic generality). With regard to the United States observation for 1900, Easterlin's data should help in explaining the second apparent contradiction to the above generalization. It seems likely that the 1900 observation lies within a pivotal era in United States history. Before the turn of the century, labor force participation rates tend to be higher in the south. After 1900 and up to 1960 this pattern reverses itself and becomes consistent with the results outlined in Table 7; the poorer Southern states are then typified by low participation rates relative to the North and West.

Columns 5 and 6 represent the results of decomposing regional income into a number of economic sectors and an examination of two of these sectors, agriculture and manufacturing. In these columns our  $V_w$  measures the degree of regional inequality where agricultural (or industrial) productivity differentials are weighted by the regional share of the agricultural (or industrial) labor force in the national agricultural (or industrial) labor force. Is regional dualism more prevalent in a traditional sector, agriculture, and one in which technology is more localized by regional resource endowment?

The answer to this question is most definitely in the affirmative, although we base it on a very limited sample, because of the rare appearance of regional income data with sector breakdown. The computed ratio of agricultural  $V_w$  to industrial  $V_w$  is given in column 8. In six out of the eight cases, regional dualism is much more severe in agriculture. The most extreme examples are Yugoslavia, Spain, Brazil, and the United States (1900), where regional income

63. See Section II for a brief theoretical discussion of the selective nature of migration.

TABLE 7.  
Labor Participation and Sectoral Distribution: By Regions

| Country<br>(1) | Year<br>(2) | Income/<br>capita<br>(Vw)<br>(3) | Income<br>or prod. /<br>worker (Vw)<br>(4) | Agric. prod. /<br>agric. lab.<br>(Vw)<br>(5) | Indus. prod. /<br>Indus. lab.<br>(Vw)<br>(6) | (3) ÷ (4)<br>(7) | (5) ÷ (6)<br>(Vw)<br>(8) | (5) ÷ (6)<br>(Vw)<br>(9) |
|----------------|-------------|----------------------------------|--|--|--|------------------|--------------------------|--------------------------|
| Australia      | 1954/55     | 0.055                            | 0.024                                      |  |  | 2.292            |                          |                          |
| Spain          | 1957        | 0.387                            | 0.304                                      | 0.373  | 0.148  | 1.273            | 2.520                    | 3.949                    |
| Brazil         | 1950        | 0.732                            | 0.571                                      | 0.578  | 0.297  | 1.282            | 1.946                    | 2.424                    |
| Italy          | 1951        | 0.363                            | 0.321                                      |  |  | 1.131            |                          |                          |
|                | 1960        | 0.372                            | 0.303                                      | 0.357  | 0.227  | 1.228            | 1.573                    | 1.334                    |
| Japan          | 1959        | 0.259                            | 0.372                                      | 0.177  | 0.283  | 0.696            | 0.625                    | 0.573                    |
| Finland        | 1958        | 0.313                            | 0.228                                      | 0.147  | 0.141  | 1.373            | 1.043                    | 0.994                    |
| Sweden         | 1944        | 0.311                            | 0.213                                      |  |  | 1.460            |                          |                          |
|                | 1960        | 0.192                            | 0.133                                      |  |  | 1.444            |                          |                          |
| Yugoslavia     | 1959        | 0.332                            | 0.103                                      | 0.470  | 0.160  | 3.233*           | 2.938                    | 2.573                    |
| United States  | 1900        | 0.322                            | 0.384                                      | 0.461  | 0.160  | 0.839            | 2.881                    | 1.941                    |
| France         | 1951        | 0.327                            | 0.285                                      | 0.331  | 0.254  | 1.147            | 1.303                    | 1.398                    |
| Canada         | 1931        | 0.272                            | 0.272                                      |  |  | 1.000            |                          |                          |
|                | 1951        | 0.206                            | 0.179                                      |  |  | 1.151            |                          |                          |
| Colombia       | 1951        | 0.604                            | 0.568                                      |  |  | 1.063            |                          |                          |
| Austria        | 1957        | 0.225                            | 0.194                                      |  |  | 1.160            |                          |                          |

Sources (refers only to data not used in previous tables):

- (1) Spain. The data used to derive columns 4, 5, and 6 are taken from Banco de Bilbao, *Renta Nacional de España y su Distribución Provincial, 1957* (Bilbao, 1958), pp. 20-21 and 46-47. The regional units are the same as in Table 1.
- (2) Brazil. *Revista Brasileira de Economia*, Ano 14, No. 1 (March 1960), p. 119; and *Anuário Estatístico de Brasil, 1960*. See Table 1 for regional units.
- (3) Italy (1951). Tagliacarne's estimates in *Moneta e Credito* (December 1961), pp. 81-84; and Svimez, *Cento Anni di Statistiche sulle Regioni D'Italia* (Roma, 1961), p. 22. (1960) Tagliacarne's 1960 estimates, pp. 48-50 and 44-46. See Table 1 for regional units.
- (4) Japan. see Table 1.

- (5) Finland. Finland's officiella statistik, *Inkomst-och Förmögenhets-statistik*, 1958 (Helsinki, 1961), Table 2, pp 52-53; based on sixteen "economic regions."
- (6) Sweden. (1944) Uses 1940 participation rates derived from Statistiska Centralbyrån, *Statistisk Årsbok för Sverige*, 1945 (Stockholm, 1945), T-27, pp. 36-37. (1960) Statistiska Centralbyrån, *Skattelastningarna Samt Fördelningen av Inkomst och Förmögenhet, 1961* (Stockholm, 1962), T-18, p. 48. See Table 1 for regional units.
- (7) Yugoslavia. The figure in Column 4 is very suspicious. The data was taken from *Statistički godišnjak FNRJ, 1961* (Beograd, 1961), pp. 316 and 350; based on eight provinces. The reader should note that it is possible for both columns 5 and 6 to exceed column 4, since we have not examined the service industry.
- (8) United States. Derived from Easterlin, in *Trends*, *op. cit.*
- (9) France. Derived from *Études et Conjuncture*, Supplement (1955), pp. 18-19 and 85-87; see Table 1 for regional units.
- (10) Canada. (1931, 1951) Dominion Bureau of Statistics, *National Accounts: Income and Expenditure, 1926-1956* (Ottawa, 1958), T-28, pp. 64-65, and Appendix T-1, pp. 100-101; R. D. Howland, *Some Regional Aspects of Canada's Economic Development* (Ottawa, 1957), p. 78. See Table 1 for regional units.
- (11) Colombia. Uses 1951 population weights and labor force estimates, but 1953 income estimates. *Estudio Sobre las Condiciones del Desarrollo de Colombia*, Misión Economía y Humanismo (Bogotá, 1958), pp. 19 and 326.
- (12) Austria. Uses 1951 population and labor force estimates. Österreichischen Statistischen Zentralamt, *Statistisches Handbuch für die Republik Österreich, 1958* (Wien, 1958), p. 10. See Table 1 for regional units.

## REGIONAL INEQUALITY AND NATIONAL DEVELOPMENT 199

inequality in the agricultural sector is approximately two to two-and-a-half times that of industry.<sup>64</sup> In postwar Finland, regional inequality in agriculture is only slightly greater than that of industry, while, again, Japan is the exception—regional dualism in industrial production is more severe.

Our conclusions are not significantly altered if we compare unweighted indices of sectoral inequality. The ratio of regional productivity variation in agriculture to that of industry, without weighting, is given in column 9. Here again, regional dualism is more striking in the traditional sector, where regional natural resource endowment plays a significant role. At the risk of oversimplification, it appears that the persistence of high degrees of regional income disparities in such countries as Spain, Brazil, Italy, Yugoslavia, and the United States (at the turn of the century) can be further decomposed into two parts: (1) tremendous differentials in agricultural productivity, and (2) significant regional differences in economic structure (the relative importance of manufacturing employment). It would appear that regional "dualism" in the industrial sector plays a minor role, and that its significance has been grossly exaggerated in much of the current development literature.

### IX. Summary

This concludes our investigation into the nature of regional dualism. What we have done thus far is to simply describe the nature of the so-called "North-South problem," giving particular attention to the relationship between regional dualism and national economic development. There is a consistent relationship between the two: rising regional income disparities and increasing North-South dualism is typical of early development stages, while regional convergence and a disappearance of severe North-South problems is typical of the more mature stages of national growth and development.

More specifically, both our cross-section approaches and our time series analysis suggest that there is a systematic relation between national development levels and regional inequality or geographic dispersion. In the international cross-section, the degree of regional inequality is very high in Europe's middle income class, but consistently lower as we move to higher levels of development. Although our evidence is much less extensive, it also appears from this sample that those nations below the middle income class have not yet generated the high levels of regional inequality associated with Spain, Italy, Colombia, and Brazil. The U.S. cross-section lends support to the international cross-section, in that the states with lowest income per capita are also typically those with the greatest inter-county inequality. The historical evidence on regional productivity or income per capita differentials is much more difficult to collect, but what little information we have on 19th and 20th century Italian, Brazilian, U.S., Canadian, German, Swedish, and French experience suggests that increasing regional inequality is generated during the early development stages, while mature growth has produced regional convergence or a reduction in differentials. Finally, we have seen that regional dualism or inequality is much more extensive within the agricultural than within the industrial sector, and that later participation rates in past contribute to regional income per capita differentials.

64. Based on somewhat different information, this generalization seems to hold historically for France, Germany, and the United States. The ratio of  $V_{ag}$  to  $V_{ind}$  in agriculture to that of industry ranges between 1.1 and 1.9 over the period 1850-1920.

This leaves us with a number of interesting related questions which are left unanswered in this study. The most pressing question is, of course, why does this pattern of regional inequality persist? What is the mechanism by which regional income differentials increase in early development stages, then stabilize, and then diminish in mature periods of growth? Have economic institutions in the past been of such a nature to cause capital to first flow in an interregional fashion, so as to increase the income gap between North and South, and then to cause this flow to reverse? Are presently developing nations sufficiently aware of the conflicts between national growth optimization and regional equality? If they are, are they aware of the costs necessary to reduce such inequities in early development stages? What historical role have central governments played in contributing to these patterns of regional inequality, and can contemporary developing nations derive benefit from that knowledge? What role do changing patterns in regional income distribution play in contributing to changes in national size distribution?<sup>65</sup>

But the most important question, one which is related to those enumerated above, has not yet been posed. If, indeed, contemporary underdeveloped nations are attempting to achieve industrialization on a weaker and more unstable socio-political scaffolding, "can... the underdeveloped societies withstand the strain which further widening of income inequality is likely to generate?"<sup>66</sup>

These questions seem extremely important. Hopefully, economists will continue to find them interesting enough so that some answers will appear in future research.

65. See Eugene Smolensky, "Industrialization and Income Inequality: Recent United States Experience," *Regional Science Association Papers*, VII (1961), 67-88. Some historians have even suggested the use of regional income inequality indices to approximate the historical patterns of national size distribution!

Kuznets, of course, has emphasized the importance of intersectoral distribution as a contributor to size distribution trends. Given the information here contained on regional inequality, is it possible that *changes* in national size distribution are dominated by a combination of changing regional differences within sectors and changes between sectors?

66. Kuznets, "Economic Growth and Income Inequality," *op. cit.*, p. 26.

## 8 Internal and External Factors in the Development of Urban Economies

Wilbur R. Thompson

Urban-regional growth theory has subsisted since the uncertain date of initial systematic enquiry, sometime in the early twenties, on one simple but powerful idea, the export base concept. Because the history of this concept has been traced at length elsewhere,<sup>1</sup> as a prelude to the present discussion we need only remind ourselves of the concept's main characteristics and limitations as brought out through the years of debate.

### THE LEGACY OF THE EXPORT BASE THEORY

From the beginning, economic geographers, economic historians, urban planners, and urban land economists have all seen the usefulness of distinguishing between those local industries which sell outside the "local economy" (later interpreted more precisely as the job commuting radius about the urban center) and those which sell within the "local labor market." The export industries clearly generate a net flow of income into the local economy from which the necessary imports can be financed. In this most immediate sense of current money flow, the export sector is basic and the local service sector is derivative in origin.

When, in the postwar period, economists began to take a more sustained interest in regional development, a parallel was seen between the export base concept and the export multiplier concept from Keynesian economics. The contributions of the economists here tended to reinforce the notion of the primacy of the export sector in the performance of the small-area, open economy because usually the economist was thinking in cyclical terms—of the expansion and contraction of output and employment within given export plant facilities. The direction of causation is clear as national cyclical change ebbs and flows through the portal of the local export sector and swirls about the passive stores and offices.

But the attempted marriage of the export multiplier and the export base theory was more than a little forced. Those with deeper and more sustained involvement with the city—urban planners especially—did

<sup>1</sup> See especially the series of nine articles by Richard B. Andrews, in *Land Economics*, Vol. 29, No. 2 (May 1953)—Vol. 32, No. 1 (February 1956), and the writings of other early proponents and critics of the export base theory of regional growth, all reprinted in Ralph W. Pfouts, *The Techniques of Urban Economic Analysis* (Chandler-Davis, 1960).

Reprinted from *Issues in Urban Economics*, ed. Harvey S. Perloff and Lowdon Wingo, Jr. (Baltimore: Johns Hopkins Press [for Resources for the Future, Inc.], 1968), pp. 43–62, with permission of the author and The Johns Hopkins Press. © 1968 by The Johns Hopkins Press.

not have in mind the direction of causation of variations in per capita income of a (nearly) fixed population and labor force over the business cycle. Urban planners care more about the lines of linkage in aggregate growth of total population over decades. Consequently they were interested in a very different kind of "multiplier" that defined locational linkages in the long run where new investment in plant and plant relocations attract other new plants and homes and stores.

Through the postwar period it has not always been clear when a given discussion was addressed to cyclical changes in local output, income, and employment of interest to, say, local public revenue and public assistance officials, or when long-run public investments in streets, utilities, schools or whatever were at issue. Both the projecting of next year's tax revenue and the pouring of cement are significant social questions, but the time horizons are, of course, quite distinct, as is also the relevant theory.

Understandably, some urban planners rebelled at the invidious comparison suggested by the terms "basic" and "local service," especially as these words implied a higher policy priority to the needs of the export sector. The export firms were more often than not those drab manufacturing plants on the edge of town, or currently moving there, while the local service sector was clustered in the more elegant high-rise buildings that dominated their favored downtown sites and symbolized "city" itself. Usually, the urban planner was not sophisticated enough in economics to defend well his intuition that in his world of the very long run (locational economics) it may well be the local service sector which is enduring (basic) and the manufacturing plant which is transitory.

As we move from local cycles to local development, we shift our interests from money flows, which clearly run from exports to local services, to concern for comparative costs relative to competing urban centers. The lines of causation are then turned inside out to the degree that comparative costs rest on the efficiency of local transportation systems, public utilities, banks and other financial institutions, schools and universities and other training and retraining centers, and a host of other critical supporting services. Such services are, moreover, likely to be more permanent than the particular factories of a given time period.

We begin to see more clearly the relationship between the time period and the lines of causation in urban change and development when we review more precisely the simplistic and somewhat tenuous dichotomy of export and local service. With time and the contributions of the economist, the concept of the export sector has been extended beyond direct exports to other regions to include local firms which supply



intermediate products to local exporters. As one moves, however, from the more obvious "linked industries" (tire manufacturers in Detroit, for example) farther and farther back in the stages of production to trucking firms hauling inputs and outputs, to accounting, financial and promotional services, and even back to technical curricula in local schools, a near continuum of activity emerges. It becomes very difficult to dichotomize local activity, and the most basic work is indeed performed by the very same banks, schools, and utilities that service the households.

These complex lines of linkage among the many local activities can be, and were, brought out crisply and definitively in the many local input-output tables constructed around the country. But, again, the input-output tables quantify the flows among a given set of local industries and are most relevant to short-period variations, although some limited extension to growth paths is possible. We do not, however, have the full development equivalent, a locational matrix that tells us which industries follow a given industry to a locality and when. This is what the urban planners and economic geographers had in mind.

With greater urban size comes a tendency toward greater local self-sufficiency. More than a decade ago, critics of the early, more simplistic form of the export base theory noted the tendency for the base-to-service ratio to fall with increased size. While perhaps none of the early developers of the export base mechanism would have insisted, on being pressed, that the ratio of export to local service workers was fixed, they all too easily slipped into the habit of multiplying the change in the number of, say, manufacturing workers by some customary fixed coefficient to find the change in total employment and then remultiplying that employment figure by the (inverse of the) labor force participation rate to yield the change in total population of a community. Perhaps the number of retail food store clerks or barbers per thousand population is nearly constant across the spectrum of urban areas, but with larger size the threshold of economic local production for the local market is passed for one new activity after another. First come those activities with only modest economies of scale and high transportation costs (medical clinics, for example), and with great size even those activities with very large scale economies and low transportation costs (such as investment banking) come to be produced locally for local sale.

To hold the prime role for an export sector that accounts for a half of all local activity in an urban place of 10,000 population is a far cry from retaining that role when metropolitan areas reach one-half million and over and where the export sector may account for only one-quarter



of all activity. (And this applies even in cyclical analysis, where shifts in local investment and in the "local-consumption-of-local-production" function probably come to rival export multipliers in local cyclical impact.) Clearly, then, the relative roles of the export and local service sectors change with the relevant time period and the size of the urban region under analysis.

In the analyses to follow an all-embracing term will be used: the "local industry-mix." In static or very short-period analysis of a small urban area the distinctive and determining part of the local industry-mix is the current export mix, and the impact of the current local industrial specialties on local economic welfare will be traced through—on the pattern of income as well as the growth rate. But as we move through time and to larger size, the analysis will be directed to the process by which structural change takes place in the local industry-mix. The analyses will, therefore, begin with the export industry-mix and progress toward increasing involvement with the local service sector, especially those parts of it which facilitate adaptation to change and/or produce change itself.

#### URBAN-REGIONAL INCOME ANALYSIS

The easiest place for the economist to take hold of urban-regional growth and development is in the middle of the story, that is, at the present. As we come to understand better how a given—the current—industrial structure affects the performance of a local economy, even in static terms, we will come to see better the inexorable development of forces leading to change. The most conventional and powerful tools and skills of the economist can, moreover, be most quickly brought to bear in a static regional income analysis. Here we can begin simplistically by casting a local economy as a mere bundle of industries in space: Tell me your industries and I will tell your (immediate) fortune. How could a highly specialized economy fail to reflect in its level, distribution, and stability of income and growth rates its distinctive industry mix? Our analysis will, moreover, retain those traditional dimensions of economic welfare that follow from national income analysis.

The major point of departure will be that no attempt will be made to discuss the general level of income or the general degree of inequality or over-all national economic stability; only the pattern of regional deviations about the national average. Our charge in urban-regional economics is to determine why some urban areas are richer or more egalitarian or more unstable—or all three (as in the case of Flint, Michigan). Urban-regional economics, freed in the beginning from

explaining the general state of the national economy, becomes the study of variations within the system of cities.<sup>2</sup>

While urban-regional income analysis is more than merely industrial analysis, it is intriguing and surprising to see how far one can go in explaining the level, distribution, and stability of local income as mere extensions of the local industry-mix. But because this has been done elsewhere at considerable length,<sup>3</sup> it should be sufficient here to summarize what seems, deductively, to be the more important lines of interaction between the local industrial structure and the local income pattern before pushing beyond into the much more difficult subject of change.

#### *The Level of Income*

The many subtle and interacting forces that operate to enrich a community may be factored into two generic determinants of the level of income: skill and power. We would expect, typically, to find above-average skill in local labor forces that are wrestling with the production of the newer products. Implicit here is some form of the "learning curve": new tasks tend to be the most demanding of intelligence and creativity, but in time the production process becomes rationalized and relatively routine. Not only are skilled workers more scarce and therefore higher paid, but skilled workers on new products tend to share the monopoly power of the early lead in a new industry; the high profit margins enjoyed, temporarily, by local innovators are passed on in part to local labor. Innovation, then, gives rise to both the skill needed to rationalize the new and strange and the price power of an early lead.

Other local economies producing older, slower-changing products may also profit from the weak competition that originates in pronounced internal economies of scale relative to the size of the market. Large investments in plant and equipment give rise to high fixed costs; the latter leads to substantially lower unit costs at very large outputs; this leads, in turn, to fewer plants (firms) and (implicit) collusion in pricing. Large investments reinforce this market power by impeding the entry of new firms as potential competitors. The price power of oligopoly is not enough, however, to enrich a community, especially if

<sup>2</sup> Worthy of attention at some other time and place is a careful investigation of the reverse side of the coin: the way in which a non-optimal system of cities depresses the national income.

<sup>3</sup> Wilbur R. Thompson, *A Preface to Urban Economics*, 1965, Chaps. 2-5; or for a revised but briefer version, Thompson, "Urban Economic Development, in Werner Z. Hirsch (ed.), *Regional Accounts for Policy Decisions*, 1966 (both books, The Johns Hopkins Press, for Resources for the Future).

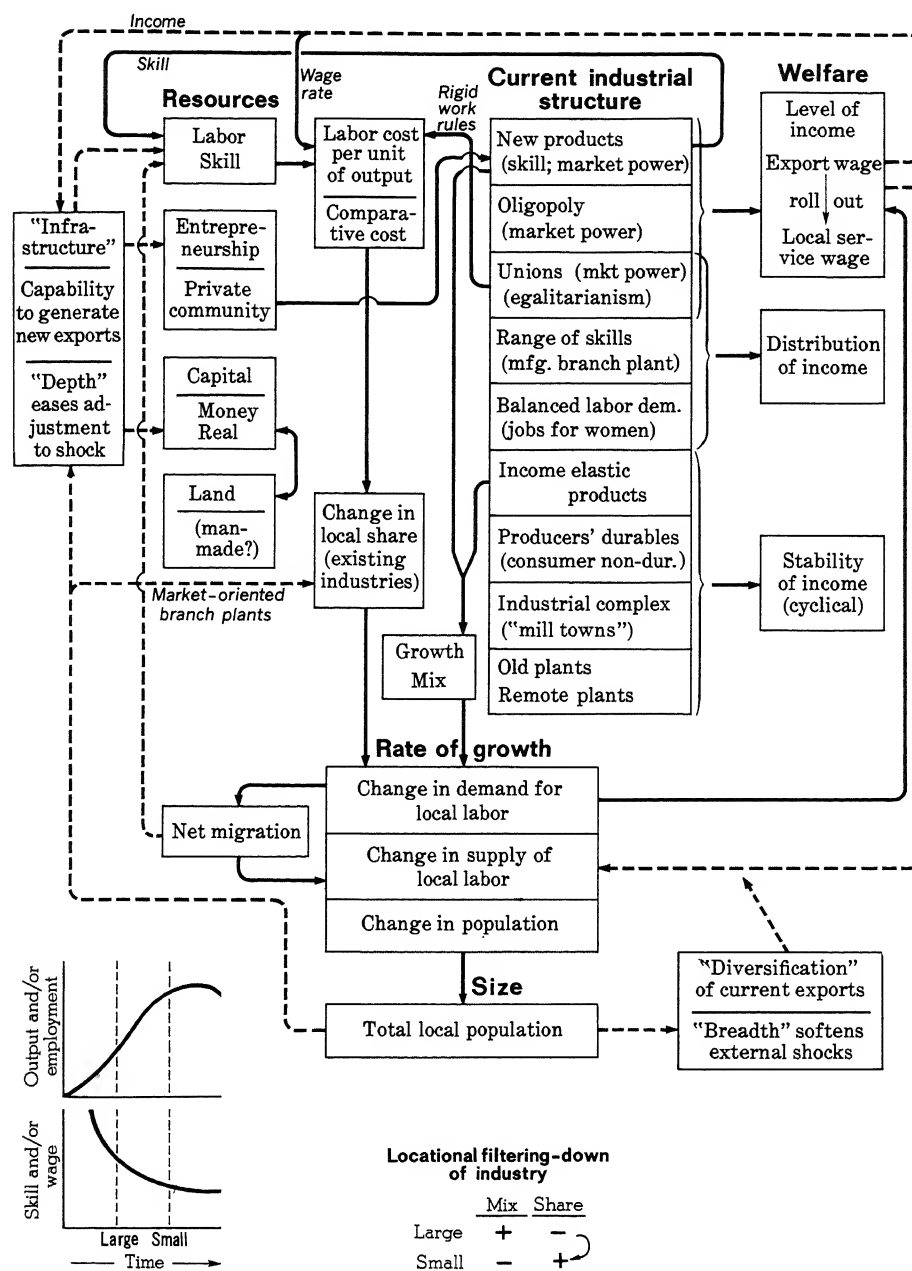
the local plant is absentee-owned. Where there are large, multiplant corporations whose stockholders are scattered, oligopoly power in the local export sector can be translated into high wage rates and/or family incomes only when it is wedded to labor power. The necessary condition would seem to be an aggressive, industry-wide union only nominally resisted by a secure oligopoly that is able to support higher wages through higher prices.

Even a moderate degree of price elasticity in the demand for local labor in the oligopolistic export industry is not likely to constrain local income if the adverse employment effects of high wage rates can be worked off in an expanding market (in the form, perhaps, of a new or income-elastic product) or by easy out-migration. If the oligopolistic export industry is a growth industry, the community will enjoy the benefits of economic power to the fullest as the monopoly wage rate of the export sector "rolls-out" into the local service sector, enriching local teachers, bus drivers, and bankers by more than it deflates the money income of the export workers through increased local prices. Wage roll-out assumes that there is some appreciable labor mobility between the export and import sectors and some appreciable immobility of movement into the community from the outside, so that the high export wage largely benefits local residents. The effects of innovation, market power, and unionization on the level of income are summarized in Figure 1.

#### *The Distribution of Income*

Unions not only raise local wage rates and average family incomes; they also tend to reduce the degree of inequality. The frequent threat of skilled workers to break away and create their own union is indirect evidence either that "unionism" (ideology) or economic democracy (assuming majority rule and representative union government) is more egalitarian than the free competitive market. The lesser degree of (median family) income inequality exhibited by the more highly industrialized metropolitan areas is probably due in significant measure to the relatively narrow range of skills characteristic of manufacturing, especially branch plant, assembly operations. Still, the labor union would seem, a priori, to be an equalitarian force of comparable magnitude.

Some appreciation of the complexity of the forces determining the distribution of local income is gained by noting that the local labor markets that generate a large number of jobs for women tend to be among the most egalitarian of all. A high female labor force participation rate may come out of light manufacturing (as in a textile town),



but probably more often it is the commercial, financial, medical, educational, or governmental center which generates a balanced demand for labor.<sup>4</sup> Implicit here is the assumption, for which there is reasonably good supporting evidence, that working wives and other female second-income earners come more than proportionately from the lower-income households. Certainly, second-income earners also raise the average family income but their impact on the distribution of family income would seem to be more significant both quantitatively and socially.

#### *The Stability of Income and Employment*

Two standard concepts from "national economics" seem highly transferable in any attempt to trace through the lines of linkage between the local industrial structure and the relative severity of the local business cycle. Those areas specializing in producers' durable goods should be most unstable and those specializing in consumer non-durables most stable, following from the greater variation in investment expenditures than in consumer expenditures over the cycle and from the simple fact that durable goods can easily be made to last a little longer while the replacement of non-durables is less postponable. The early statistical evidence suggests, however, that an industry-mix explanation of the local cycle will not be sufficient.

Specialization in durables is not, for example, significantly correlated with the degree or rate of decline in employment during the 1957-58 recession. True, the census data do not permit sure and easy separation of *producers' durables* from consumer durables. Still, the tendency for durable goods to cluster (agglomerate in industrial complexes) while non-durables often stand alone as the sole support of the smaller place (the textile town, for example), may blur the sharp national contrast by posing industrial centers concentrating on *diversified* durable-goods against isolated towns concentrating on *specialized* non-durables. Any significant lead or lag in the timing of the industry cycle of any of the local durable industries softens the aggregate local fluctuation as the leading industry begins to fall while other local industries are still climbing, and begins to rise from its trough while the others are still falling.

Further, as we translate national economic concepts into regional economics, we must remember that a local economy is not only an atypical bundle of industries, it tends also to be a non-random, or a

<sup>4</sup> Since these non-manufacturing places tend to create more income inequality out of their wider ranges of occupations, their higher propensity to produce jobs for women tends only to make them less unequal than they would otherwise be, but still not as egalitarian as factory towns.

biased, sample of plants by age. The older industrial areas tend to have more of the older plants, and these tend to be the higher-cost facilities that are cut back first and brought back last over the cycle. To the degree, moreover, that the newer plants are more capital-intensive (automated), the higher ratio of fixed to variable (labor) cost should lower marginal cost and reinforce the tendency for the newer facilities to absorb more than their share of the smaller recession market and lesser share of the larger boom output.

Let us draw out the asymmetry of national and regional cycle analysis a little more by adding that most neglected facet: space. A locality may have relatively high-cost facilities not because of age but because of a more out-of-the-way location. Rarely do we incorporate the cost of transportation of either inputs or outputs into our cost curves. To do so, however, is to see clearly that remoteness from the market raises marginal delivered costs and amplifies output variations in response to demand and price fluctuations. Moreover, "remoteness" may have to be interpreted in the context of differential oscillation of regional markets. Products with high transportation cost are not ordinarily sold across the nation. Thus, a locality's cycle will tend to reflect the regional industry-mix as well as its own mix and its locational position in the region. There is little doubt that a paper-mill town in a region with a state capital and a large university would be in relatively good shape over the business cycle.

#### *Growth and Income Patterns*

By influencing the rate of growth of the local economy, local industrial characteristics also act indirectly to shape the local income pattern. A local export sector which emphasizes either new products or income-elastic products will tend to experience a greater than average expansion in output and in demand for labor. A steadily rising national per capita income acts directly to stimulate the growth of localities producing income-elastic products. Through increasing discretionary income more than proportionately, growing affluence also acts indirectly to stimulate the demand for new products. The buoyancy of a new-products economy comes not merely from the opportunity to exploit a new market unconstrained by depreciation rates and replacement schedules, but also from the eager search for variety—temporary enchantment with a new product becomes a persisting obsession for new products.

A greater than average increase in the demand for local labor interacts with an average increase in the supply of local labor (typical birth, death, and labor force participation rates) to create rising wage rates, overtime, upgrading, part-time jobs, relaxed job requirements,

and other reflections of a tight labor market. A greater than average rate of growth in the demand for local labor should then lead to a relatively fast rising and, ultimately, a relatively high median family income, but even more surely should lead to a lessening of the degree of local income inequality. The marginal worker who finds a full-time job will, of course, experience the greatest rate of increase in income. Recent studies of income inequality have shown that it was the tight labor market of the war years which effected the great decrease in inequality and that there has been little or no reduction in interpersonal inequality since the end of World War II.<sup>5</sup>

Again, we distinguish between the national economy, which is a closed economy, and the local economy, which is open to the flow of people as well as goods. Classical economic theory would predict an increase in the supply of labor as the high wage rates, abundant jobs, and high welfare payments of this prosperous community attract the unemployed and unemployable. The seeming paradox of poverty amidst plenty in the booming locality is quite in conformance with classical theory. In the long run, growth industries do not make a locality richer, only bigger.

## TOWARD A THEORY OF URBAN-REGIONAL GROWTH

### *Population Size*

Since size is simply the cumulation of growth, those places which grow faster tend to get big more quickly and are at any given time larger. Growth creates size and size reacts to restructure the local economy so as virtually to ensure further growth at a near average rate—reacts, that is, to produce growth stability. The simplest, most dramatic and thus most widely appreciated structural change accompanying large size is toward a diversified mix of current exports. A large sample of exports, such as we find in Chicago or Philadelphia, is likely to include some very new, fast-growing industries, some middle-aged slow-growing ones, and even a few that are old and declining. Moreover, a large sample would be likely to mix income-elastic with income-inelastic products. Consequently, we would not expect to find the large, diversified export sector generating unusually rapid growth or very slow growth.<sup>6</sup>

<sup>5</sup> See for example, Herman P. Miller, *Income of the American People* (Wiley, 1955).

<sup>6</sup> The coefficient of variation ( $\sigma/\bar{x}$ ) of the per cent change in population, 1950-60, for the twenty-two metropolitan areas with a million population or more was about one-half that of the smaller metropolitan areas. See Thompson, *A Preface to Urban Economics*, p. 192; or for a more extended treatment of

The very large metropolitan area is even more distinctive in its depth than in its breadth. The local social overhead—the infrastructure—that has been amassed is, more than export diversification, the source of local vitality and endurance. Stable growth over short periods of time, say up to a decade, is largely a matter of the number of different current exports on which employment and income are based. But all products wax and wane, and so the long-range viability of any area must rest ultimately on its capacity to invent and/or innovate or otherwise acquire new export bases.

The economic base of the larger metropolitan area is, then, the creativity of its universities and research parks, the sophistication of its engineering firms and financial institutions, the persuasiveness of its public relations and advertising agencies, the flexibility of its transportation networks and utility systems, and all the other dimensions of infrastructure that facilitate the quick and orderly transfer from old dying bases to new growing ones. A diversified set of current exports—“breadth”—softens the shock of exogenous change, while a rich infrastructure—“depth”—facilitates the adjustment to change by providing the socioeconomic institutions and physical facilities needed to initiate new enterprises, transfer capital from old to new forms, and retrain labor.

#### *Impact of Size on Local Resources*

Whatever may be the prime source of national economic development, entrepreneurship must be the heart of comparative regional growth in any system of open regions. The large urban area would seem to have a great advantage in the critical functions of invention, innovation, promotion, and rationalization of the new. The stabilization and even institutionalization of entrepreneurship may be the principal strength of the large urban area. In an earlier work, the argument was advanced that a large population operated to ensure a steady flow of gifted persons, native to the area. A population of 50,000 that gives birth to, say, only one commercial or industrial genius every decade might get caught between geniuses at a time of great economic trial such as the loss of a large employer, but in a population cluster of 5 million, with an average flow of ten per year, a serious and prolonged crisis in local economic leadership seems highly improbable.<sup>7</sup>

---

the effect of great size on stability, see Wilbur R. Thompson, “The Future of the Detroit Metropolitan Area,” in W. Haber, W. A. Spivey, and M. R. Warshaw (eds.), *Michigan in the 1970's: An Economic Forecast* (University of Michigan, 1965).

<sup>7</sup> Thompson, *A Preface to Urban Economics*.



On further thought, the great viability of the large aggregate in a time of quick and sharp change seems to lie even more in the separate but co-ordinated institutionalization of the many entrepreneurial functions, especially invention, innovation, and quick adaptation to change. The very large metropolitan area typically hosts a large state university with well-developed programs in basic research and in graduate, professional, and continuing education. The main medical centers grow up to serve the nearby population, and the many advantages of scale draw medical research personnel and funds. As we become more of a service-oriented economy, the city itself becomes the very product that is being redesigned and re-engineered—becomes the experiment as well as the laboratory. Small wonder that the largest metropolitan areas can be so little concerned with promoting area industrial development, compared with the frantic activities of this kind conducted by the smaller areas.

When we turn to consider the co-operating factors of production, the case for the large urban area does not suffer. Money capital has long been recognized as the most spatially mobile factor, and is so handled both explicitly and implicitly in the locational analyses of Edgar Hoover and August Lösch. The advent of the modern large corporation, financing internally from depreciation reserves and retained earnings, and externally in a national money market, further weakens the locational influence of local capital supplies for the large, well-established, national-market business. But existing real capital is highly immobile—most often permanently fixed in space. The large urban area has by definition the largest and most varied supply of existing real capital—infrastructure—referred to above as the real economic base of the city. It is, in fact, difficult to define the boundaries between entrepreneurship and capital; consider the research laboratories of the large universities—a case where public capital spins off new products and new businesses.

Conversely, small business is often highly dependent on local capital supplies, both commercial credit for working capital and equity money. But the greatest proportion of small business is in local services, making such activity derivative from, not determinant of, regional growth, at least in the short run. To the extent, moreover, that local service activity does play a significant developmental role in the very long run, the large urban area is probably better able to finance this special group of small business which does depend on local capital supplies.

Turning to labor as a locational factor, the case for the large urban area is again reinforced. With the spread of unionism comes the spatially invariant wage, an indispensable concomitant of labor's bargain-

ing power. An industry-wide wage rate in a given industry is a big step toward spatially invariant labor costs, although regional differences in productivity may remain owing to regional differences in skill and motivation. But probably, on net, increasingly automated production processes operate to reduce the opportunity for variations in worker skill or effort to a point where the quantity or quality of output is not significantly altered.

When workers quote the same wage for a given occupation at all places, they give up all influence on the location of work. If blue-collar, middle-income workers should happen to prefer small towns or medium-size cities as places that offer pleasant living and fishing, this is irrelevant as a locational factor. What could be most relevant is that the wives of the corporate managers prefer the theatre. Unionism acts to shift location from a production to a consumption decision and increasingly into the hands of management—urbane professionals.

Finally, let us consider land, which has always been a critical determinant of the locational pattern of economic activity. Most urban areas began by exploiting some natural resource or strategic position on the earth's surface; but for some time now, the share of activity accounted for by the extractive industries and raw-material-oriented processing industries has been declining. However, while on the production side the tie to land has loosened, increasing affluence has tightened it on the consumption side. With high incomes and greater leisure, land in the form of natural beauty and good climate retains significant locational power. But we must also recognize that more and more we live in a "man-made environment" where we fabricate lake-front lots and ski-slope snow, and where we can probably look forward to plastic trees and mountains. That is, capital becomes increasingly a substitute for land; capital is subject to economies of scale and is the magic wand of the entrepreneur-illusionist.

#### *A Filtering-down Theory of Industry Location*

If the larger urban areas are, in fact, more than proportionately places of creative entrepreneurship, then we might hazard a broad hypothesis on the nature of regional growth patterns—one that can be tested with available data.<sup>8</sup> Following current practice, let us factor regional growth into two major components: mix and share. An area may grow rapidly either because it has blended a mix of fast-growing industries (those of new products as in Los Angeles, or those with income-elastic demands as in Detroit), or because it is acquiring a

<sup>8</sup> U.S. Department of Commerce, *Growth Patterns in Employment by County, 1940-50 and 1950-60*, Vols. 1-8 (U.S. Government Printing Office, 1965).

larger share of the older, slow-growing ones (the movement of the textile industry to various North Carolina towns is such a case). The hypothesis offered is that the larger urban areas tend to combine a fast-growing mix with a steadily declining share of these growth industries. We modify here, then, the earlier position that the large urban area is highly diversified by suggesting that its rich mix is biased at least a little toward growth industries, but that its growth is dampened by the steady spinning-off of these industries. The result is typically a near-average growth rate.

The larger urban area is believed to invent, or at least innovate, to a more than proportionate degree and, therefore, to enjoy the rapid growth rate characteristic of the early stage of an industry's life cycle—one of exploitation of a new market. As the industry matures into a replacement market, the rate of job formation in that industry slows nationally and the local rate of job formation may slow even more if the maturing industry begins to decentralize—a likely development, especially in non-unionized industries, because with maturity the production process becomes rationalized and often routine. The high wage rates of the innovating area, quite consonant with the high skills needed in the beginning stages of the learning process, become excessive when the skill requirements decline and the industry, or parts of it, “filters-down” to the smaller, less industrially sophisticated areas where the cheaper labor is now up to the lesser occupational demands.

A filter-down theory of industrial location would go far toward explaining the southern small-towns' lament that they always get the slow-growing industries. Out-of-the-way towns like these find they must run to stand still, because their industrial catches come to them only to die. Meantime, these smaller industrial novices struggle to raise per capita income over the hurdle of industries which pay the lowest wage rates. Clearly, the characteristics of slow growth and low wage rates (low skills) might be viewed as two facets of the aging industry. The smaller, less industrially advanced area struggles to achieve an average rate of growth out of enlarging shares of slow-growth industries, originating as a by-product of the area's low wage rate attraction.

The larger, more sophisticated urban economics can continue to earn high wage rates only by continually performing the more difficult work. Consequently, they must always be prepared to pick up new work in the early stages of the learning curve—inventing, innovating, rationalizing, and then spinning off the work when it becomes routine. In its early stages an industry also generates high local incomes by establishing an early lead on competition. The quasi-rents of an early lead are in part lost to the local economy, as dividends to widely dispersed stockholders,

but in part retained as high wage rates, especially if strong unions can exploit the temporarily high ability to pay. It would seem, then, that the larger industrial centers as well as the smaller areas must run to stand still (at the national average growth rate); but the larger areas do run for higher stakes.

In order to develop, it seems that the smaller, less favored urban area must attract each successive industry a little earlier in the industry's life cycle, while it still has substantial job-forming potential and, more important, while higher-skill work is required. Only by upgrading the labor force on the job and generating the higher incomes—hence the fiscal capacity—needed to finance better schools, can the area hope to break out of its underdevelopment trap. By moving up the learning curve to greater challenge and down the growth curve toward higher growth rates for a given industry, an area can encourage the tight and demanding type of local labor market that will keep the better young adults home, lure good new ones in, and upgrade the less able ones.

#### *The Delicate Balance between Equilibrating and Disequilibrating Forces*

While we have much to learn from static regional analysis, it is necessary to move toward a more dynamic framework if we are ever to understand how the intricate interactions of development arise. A fully dynamic analysis is still far off, but a form of "stages" analysis provides the means of gaining new insights into the development process. At any given time the industrial structure of a local economy acts directly on the stock of local resources so as to alter their quantity or quality. Each process or stage of regional development leaves a legacy in the form of an altered stock of labor, capital, or entrepreneurship which acts to change the path of development. Only one illustration of a possible development sequence is offered here, but it is one which brings out the delicate balance between equilibrating and disequilibrating forces in regional development.

In the static, industrial-type analysis dealt with in the previous section, it was argued that a locality could parlay an oligopolistic export industry and a strong union into high wage rates and high incomes. A locality that transforms market power into local wage rates that exceed the relative productivity of its labor force will live well but dangerously on overpriced labor. As the relevant industry matures it will slow its rate of job formation, and especially the rate of *local* job formation, if substantial decentralization takes place with maturation, as is usually the

case.<sup>9</sup> Confronted by a rate of local job formation inadequate to absorb the natural increase in the local labor force, the search for new industries is greatly hindered by its overpriced labor.

In the very long run local monopoly power may backfire in another way. Strong unions may set rigid work rules that become more binding and expensive as time passes and technology progresses. New processes which, with no greater effort than before, permit a worker to tend more spindles at one time or drill more holes in a given time period cannot be introduced locally. The new process must and will be innovated elsewhere, leaving the locality with a sharply declining share of what might still be a rapidly growing industry. In such a case, moreover, this high-cost labor market will tend also to become burdened with old plants and heavy cycles as well as slow growth. Thus the classic advantages of an early start—a pool of labor where “the skill is in the air” and bankers who know the industry and identify with it—all may be undone by rigid work rules in a world of rapid technological change.

The locality may have to go through a painful wage deflation before it can recover; in short, there may be a day of retribution. Then, again, there may not. If the beneficent oligopoly exhibits an extended stage of strong growth and is slow to decentralize, it could enrich its host area for a full generation or more, as in the case of the automobile industry and Detroit. That locality may come to acquire the education and skill which merits a relatively high local wage rate. Local affluence, however acquired, provides the private means and the tax base needed to build superior social overhead, and thereby to provide a wholly new local economic base to support new export bases. By the time the moment of truth arrives, the local labor force could have risen in education and skill to the point where it is no longer significantly overpriced, and has become, in fact, a scarce factor.

Thus, in this latter case, no equilibrating force would ever come into play. The original nexus of market power could lead instead to a set of disequilibrating forces, as power leads to affluence and affluence leads to education and skill and further affluence. Clearly, whether equilibrating or disequilibrating forces prevail is a matter of timing. The longer a locality holds market power the more likely it is that there will be no day of retribution.

Local rates of growth in employment, labor force, and population substantially different from the national average rate—or the local rate

<sup>9</sup> Even if we assume that the industry faces a nationwide union and a spatially invariant wage and, thereby, rule out relocation into lower-wage areas as operations become more routine, as argued under the filtering hypothesis above, decentralization to reduce transportation costs is a common pattern in maturing industries.

of natural increase—induce net migration flows, and thereby also create a tension between equilibrating and disequilibrating forces. Migration acts most simply to adjust the size of the local labor force and population to the level of economic activity and serves therefore as an equilibrating mechanism in the local labor market, with reference to both employment and per capita income. But migration also acts to change local population profiles to the extent that migrants differ from non-migrants. If the more mobile (migratory) persons are the younger, more talented, more energetic or aggressive, then migration may serve to reduce the future economic potential of contracting or slowly growing areas, as it drains from them the next generation of professionals and entrepreneurs. Meanwhile, the more rapidly growing areas which receive the young migrants are building up a favorable age distribution of population, ensuring future growth. Migration, then, may begin as an equilibrating force but end as a disequilibrating one, as the rich become richer and the poor become poorer.

#### *Next Steps*

Clearly, we are now enmeshed in complex questions of timing. An oligopolistic export industry may die or leave before the community it enriches has time to grow up to its inflated wage, and thus the transition to a new industrial base becomes especially difficult. Or the beneficent industry may set in place an educated second generation before it slowly fades away. Again, net out-migration may relieve the pressure of a redundant labor supply long enough to allow the locality to transfer public funds from, say, public welfare to education and be born again. Or the net out-migration may become a "brain-drain" that leaves the community worse off after each "corrective adjustment" than it was before. We would profit greatly by turning here and now to good urban-regional time series data to test some of these hypotheses and to suggest new ones. But we lack such data. Still, the available census data do permit some preliminary inquiry that can provide a beginning in what will surely be a long search into the nature and causes of urban-regional economic development. A preliminary empirical inquiry of this nature is presented in the Appendix to this paper.

#### A NOTE ON THE GROWTH OF THE LARGEST METROPOLITAN AREAS <sup>10</sup>

Any analysis of urban economic development that stresses, as this one has done, the relative cost of doing business in urban areas of

<sup>10</sup> This section results from comments made by Harold J. Barnett on an earlier version of this paper.



various size, is bound to be bullish on city size. The larger places have a clear and sizable advantage in such areas as cheaper and more flexible transportation and utility systems, better research and development facilities, a more skilled and varied labor supply, and better facilities for educating and retraining workers. Further, these economies of scale are captured by private business as lower private costs; at the same time private business is able to slough off on society various social costs that its presence imposes, such as its addition to traffic congestion and air pollution. If, then, the external diseconomies of business-created noise, dirt, congestion, and pollution are some increasing function of city size and/or density, factor market prices are biased in favor of larger urban areas and understate the true marginal costs of production in the metropolis. In the absence of sophisticated public policy and the even more sophisticated public management that would be needed to implement price reform, factor markets so biased promote urban growth and great size.

Effective limits to urban size would seem to have to originate in the household sector. Certainly, alarms of an urban crisis are almost invariably couched in the color words of amenities: congestion, pollution and other aspects of bigness, or at least poorly managed bigness. It is, in fact, not at all clear from this largely impressionistic (and frequently impassioned) literature whether the hypothesized rising costs and/or deteriorating quality of urban life with greater scale is due to some naturally scarce factor, such as fresh air or clean water, or due instead to the probable or demonstrated failure of urban public policy and management to apply the best technology and/or arrange the best combination of factors. It could be argued that urban public management is the scarce factor in the latter case, just as management has always been seen as the ultimate limit on firm size in neo-classical price theory. But the probabilities and strategies of relaxing natural resource constraints versus managerial constraints may be quite different. None of this is meant to minimize the difficulties involved in achieving better urban policy and management in our largest, most complex urban agglomerations.

To date, the most that can be made of these popular, if not classic, "problems" of great city size is that they have slowed slightly the growth of the largest urban areas: probably New York and less clearly Chicago. But, paradoxically, the loss of amenities with great size may rebound to support the growth of the second echelon of metropolitan areas. Metropolitan areas with over a million population but less than Chicago's 8 million offer substantial infrastructure in support of modern business, although they do not rival New York and Chicago on this

score. But then neither are their "problems" quite as big either—it is easier, if not quite as exciting, to live in these second echelon metropolitan areas.

We could argue that New York must wrestle with a somewhat more advanced form of each of the classic urban problems, or we could argue alternatively the equivalent: that New York must face each new problem in urban management first. Thus, New York was the first to have to learn how to handle 10 million people, and must soon be first to master the problems posed by 20 million. Each successively smaller city, roughly in its rank order, has one more example from which to profit, whether the examples be good or bad. If there is a downward sloping learning curve in urban public management and the challenges of the field are a function of size, Chicago finds the path a little easier because New York has gone before, and Detroit profits from the pioneering of both. Detroit, that is, should be able to offer in 1970 a better organized version of the 5-million population cluster, as a partial offset to the disadvantages it suffers living in the shadow of the greater choice and urbanity of Chicago.

A recent fifty-year projection of the Detroit metropolitan area, which assumed a zero rate of net migration, was criticized as being too optimistic because "the automobile industry is a maturing industry and will, of course, experience the typical slowing rate of growth." Certainly, the automobile industry will probably trace out its own variant of the classic growth curve. But to project the *long run* growth of an urban area from its current industry-mix would be to project the slowing growth and ultimate decline of all areas, as all existing industries are slowly dying. Who, in retrospect, would correlate the growth patterns from 1900 to 1950 of citrus fruits and Los Angeles, or buggies and Detroit? Why, then, must we link automobiles and Detroit for the next fifty years? Ten-year regional growth projections based on the local industry-mix may be defended, but for periods of two or more decades a static, fixed industry-mix technique is heroic if not naive.

Given the embryonic state of urban-regional growth theory, what fifty-year projection might we make for the Detroit metropolitan area? I would argue that Detroit is big enough to assemble the infrastructure necessary to: (a) support a rapidly advancing technology, (b) employ productively and creatively an increasingly educated labor force, and (c) serve an increasingly affluent and sophisticated set of households. From such a base, Detroit should grow at about an average ( $\approx$  zero net migration) rate. At one end of the spectrum, the smaller metropolitan areas should about hold their current share of population in that the population of rural areas and small towns will soon be reduced to a



bare minimum and the rural to urban migration cannot much longer feed these cities. At the other end, few would argue that New York and/or Chicago is likely to grow faster than its natural rate of increase and drain off growth from the second echelon areas. A zero net migration projection would, in fact, be an optimistic projection for either New York or Chicago, and even this high projection would not encroach on the future population growth of the Detroit economy.

Detroit's competition in the national system of cities would seem, then, to come largely from its peer group: Cleveland, Pittsburgh, Cincinnati, Milwaukee, and St. Louis. The question can be turned around from "Why project a zero net migration rate for Detroit?" to "Why not?" What cities will suck growth from these second echelon areas: the bigger ones with more problems? the smaller ones with less infrastructure? A good case can be made that each of the two dozen or so urban areas with a million to 5 million population will net out to about an average growth rate over the next fifty years, and more than double in size; New York and Chicago will pave the way, perhaps at a slowing rate (with therefore some convergence in size at the top). All this assumes, of course, the absence of national policy that would restrict the continued growth of big cities. And at this time and vantage point, it seems likely that our national policy will be directed more toward mastering the management of large population clusters than toward preventing their growth.

## 9 Po Gr Pet

### Introducti

A century tendency (try." Hist  
teenth-cen  
social ord  
land bridg  
settlement  
United St

Now, as  
connected  
characteris  
oped resou  
enterprises  
on migrati  
Without a  
are dwindl  
intolerably

Migratic  
prevented l  
al deficienc  
riers in the  
left the rur

<sup>1</sup>Joseph Kenn  
ington, D.C.,  
Some Data  
America," *Je*  
<sup>2</sup>Everett S. La  
Union of Bio  
(Washington

This chapter  
Population G  
*bution and Po*  
and suppleme  
and Growth (and  
and "Migrati  
Corporation,

## 9 Population Movements and the Shape of Urban Growth: Implications for Public Policy

Peter A. Morrison

### Introduction

A century ago, the superintendent of the 1850 census took note of "the roving tendency of our people . . . incident to the peculiar condition of their country." Historians, too, document " . . . the remarkable fluidity of the nineteenth-century urban population. . . . more like a procession than a stable social order."<sup>1</sup> Indeed, since the first migrants trekked across the Aleutian and bridge thousands of years ago, migration has played a major role in the settlement and economic development of the Americas and especially the United States.

Now, as in the past, Americans continue to migrate for reasons that are connected with the workings of national economic and social systems. A characteristic of modern economies is the quick exploitation of newly developed resources or knowledge, a process that requires the abandonment of old enterprises along with the development of the new.<sup>2</sup> Such economies depend on migration continuously to alter the population growth rates of localities. Without a tradition of migration, which moves people from areas where jobs are dwindling to those where workers are needed, economic growth would be intolerably sluggish and imbalanced.

Migration is also an important vehicle of social mobility. Many people are prevented from bettering their circumstances less because of inherent personal deficiencies than because of discrimination and rigidly drawn social barriers in their home areas. The largely positive experience of blacks who have left the rural South and of members of ethnic groups who have departed from

Joseph Kennedy, *Report of the Superintendent of the Census for December 1, 1852* (Washington, D.C., 1953), p. 15. Stephan Thernstrom and Peter R. Knights, "Men in Motion: Some Data and Speculations about Urban Population Mobility in Nineteenth Century America," *Journal of Interdisciplinary History* 1 (fall 1970): 34-35.

Verrett S. Lee, "Psychological and Social Effects of Population Growth," in International Union of Biological Sciences, *Proceedings of the Scientific Program, XVII General Assembly* (Washington: National Academy of Sciences, 1971), p. 20.

This chapter is adapted from a report prepared by the author for the U.S. Commission on Population Growth and the American Future, *Research Reports*, vol. 5, *Population Distribution and Policy*, ed. Sara Mills Mazie (Washington: Government Printing Office, 1972), and supplemented with material drawn from "A Demographic Assessment of New Cities and Growth Centers as Population Redistribution Strategies," *Public Policy*, summer 1973, and "Migration from Distressed Areas: Its Meaning for Regional Policy," The Rand Corporation, R-1103, fall 1973.

city ghettos confirms the value of a freely mobile population in fostering improvements in personal status.<sup>3</sup>

Why people move, and with what effects, are the subjects of research aimed at delineating whether and how public policy should affect geographic mobility. Although no universal statements can be made about migration, research does indicate some generalized patterns. For example, people migrating long distances tend to do so chiefly for economic reasons, while those moving within a locality are typically acting on neighborhood preferences and changing requirements for housing.

Understanding the migration process depends on synthesizing findings arrived at through the exercise of several disciplines: economic, sociological, demographic, and other influences are all embedded in the migration episode.

The perspective of economics, for example, points up the human capital aspects of migration; gauges the costs and benefits of migration flows for sending and receiving areas and traces their implications for local and regional economic growth; and examines migration as a general investment problem, considering the benefits and costs of moving.

The perspective of sociology, on the other hand, points up the influence of underlying social structures on migration, and the orientations, attitudes, and values through which this influence is transmitted. Many of the forces coming to a sharp focus in specific personal situations are organized by common developmental processes: passage through stages of the life cycle, especially sharp breaks associated with passage from one stage to another, as in marriage, family expansion, entry into the labor force, divorce, or retirement; career changes that may necessitate, preclude, or otherwise pattern geographic movement. Sociology also explores how underlying value systems shape people's mobility intentions and their capacity to act on these intentions:

**Familism:** a high value placed on family living and a corresponding devotion of time and resources to family life.

**Careerism:** orientation toward upward social mobility and a corresponding disposition to engage in career-related activities, to at least partial neglect of family ties.

**Localism:** a parochial orientation implying interests that are confined to a neighborhood or community and reference to groups whose scope is local.

**Cosmopolitanism:** an ecumenical orientation implying freedom from binding ties to a locality, reference to groups whose scope is national rather than local, and interests that are not confined to a single neighborhood or community. (The cosmopolitan resides in a place but inhabits the nation.)

<sup>3</sup>Evidence on this point is reviewed in Peter A. Morrison, "Population Movements: Where the Public Interest and Private Interests Conflict," in U.S. Commission on Population Growth and the American Future, *Research Reports*, vol. 5, *Population Distribution and Policy*, ed. Sara Mills Mazie (Washington: Government Printing Office, 1972), pp. 335-352.

In this st  
"peculiar co

#### Determinant

In exami  
phases of th  
on a destin  
(Why do the  
tive to migr

Separated  
where to m  
other. Typic  
life precipit  
the labor fo  
son migrate  
before and

#### Who Moves

Given a s  
some people  
differentiate  
social and e  
utes differ  
connected w  
tus, his educ  
**Life-Cycle**  
virtually co  
other. Migr  
cording to  
highest for  
and initial c  
numbers of  
labor marke  
forming fan  
they are wel

Suppleme  
migration as  
prospective  
the greater  
pected futur  
Changes i

<sup>4</sup>Ira S. Lowry,  
Chandler, 196  
<sup>5</sup>See Gary S. 1  
29, 50; Larry  
*Political Econ*

In this study, each of these perspectives has contributed insight into the "peculiar condition" of the American experience that fosters migration.

### Determinants of Migration

In examining the determinants of migration, it is useful to distinguish two phases of the decision process: first, deciding to move, and second, deciding on a destination.<sup>4</sup> This perspective distinguishes a question about people (Why do they move?) from one about places (Why is a given locality attractive to migrants?).

Separated for analytical purposes, the decisions on whether to move and where to move appear to be determined somewhat differently from each other. Typically, a major change or event that alters the pattern of a person's life precipitates the decision to move—for example, entering or retiring from the labor force, changing jobs, family formation or dissolution. *Where* a person migrates is, in some respects, a separate decision—perhaps made long before and merely awaiting the "right time."

### Who Moves

Given a sedentary population and an inducement to leave home, typically some people go and some stay behind. The self-selection by which migrants differentiate themselves from the sedentary population reveals the underlying social and economic influences on decision-making. Certain personal attributes differentiate migrants from nonmigrants. The principal ones are those connected with the life cycle, the person's occupation and employment status, his educational attainment, and past mobility experience.

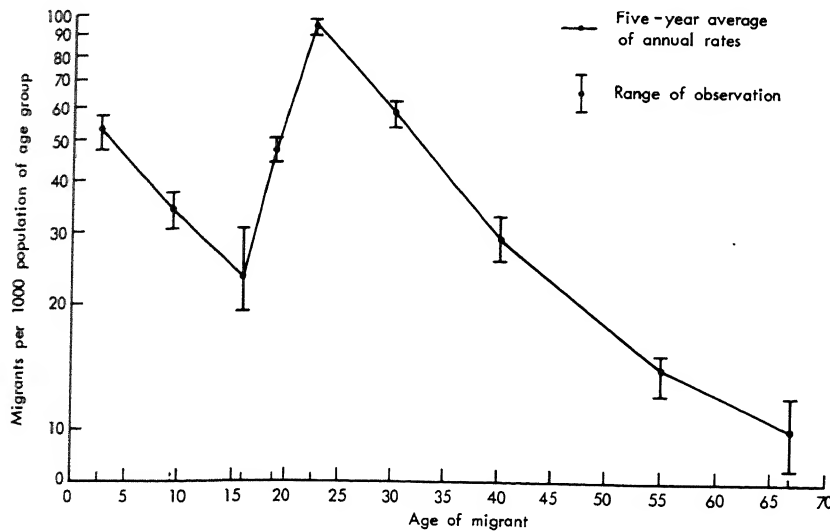
**Life-Cycle patterns in migration.** Migration is induced—in certain situations virtually compelled—by transitions from one stage of the life cycle to another. Migration rates vary sharply and with nearly universal regularity according to age groups, with young adults predominating. U.S. rates are highest for people in their early twenties, reflecting the changes, adjustments, and initial commitments that accompany early adulthood. (See fig 1.) Large numbers of these people are completing their formal schooling, entering the labor market for the first time, serving in the armed forces, and marrying and forming families. These activities predispose people to move frequently until they are well into their thirties.

Supplementing this sociological interpretation is the human capital view of migration as a form of self-investment. According to this view, the older the prospective migrant, the less his long-term expected gains from mobility and the greater its costs. The economically rational migrant maximizes his expected future earnings by moving early in his productive life.<sup>5</sup>

Changes in marital status accentuate the effect of age. Migration is closely

<sup>4</sup>Ira S. Lowry, *Migration and Metropolitan Growth: Two Analytical Models* (San Francisco: Chandler, 1966), p. 94.

<sup>5</sup>See Gary S. Becker, *Human Capital* (New York: Columbia University Press, 1964), pp. 29, 50; Larry A. Sjaastad, "The Costs and Returns of Human Migration," *Journal of Political Economy* 70 (supp., October 1962): 80-93.



Source: Calculated from U.S. Bureau of the Census, *Current Population Reports*, series P-25, various issues.

Figure 1. Annual Rates of Interstate Migration by Age of Migrant, 1965-1970

linked in time to marriage and family formation, and the first few years after marriage are a period of frequent movement. The dissolution of marriage through separation or divorce is also frequently accompanied by movement. **Occupation and employment status.** Whereas unskilled jobs are usually filled by local residents, highly skilled or specialized jobs tend to be filled through nationwide or even nationwide competition and recruitment.<sup>6</sup> Occupational differentials in migration reflect the scope of this competition and recruitment. White-collar workers migrate more frequently than manual and service workers. Within the white-collar category, professional workers (especially those who are salaried rather than self-employed) tend to be the most migratory. Professionals' migration stems from the more nearly nationwide labor markets in which such people find jobs; it also reflects the variety of relatively short-term residences that accompany higher education and internship or other apprenticeship arrangements. The least migratory occupational groups tend to be service workers and farmers and farm managers.<sup>7</sup>

Apart from these occupational differences is a consistent tendency for wage and salaried workers to move more frequently than the self-employed.<sup>8</sup>

<sup>6</sup>Donald J. Bogue, *Principles of Demography* (New York: Wiley, 1969), p. 771. See also Fred E. Katz, "Occupational Contact Networks," *Social Forces* 37 (October 1958): 52-58.

<sup>7</sup>For comparative data on these points, see U.S. Bureau of the Census, *Current Population Reports*, series P-20, no. 210 (January 15, 1971), tables 8, 9, and later issues of series P-20.

<sup>8</sup>Jack Ladinsky, "Occupational Determinants of Geographic Mobility among Professional Workers," *American Sociological Review* 32, no. 2 (1967): 253-264, and "Sources of

This differential arises from common constraints that discourage movement among the latter group. Salaried professionals are quite footloose for reasons just mentioned, but self-employed professionals typically are entrepreneurs tied down by capital investments and a clientele built up over many years. Even such specific limitations as state-based licensure for certain professions retard interstate migration.<sup>9</sup>

**Educational attainment.** Education is also related to the propensity to move. People with at least some college education, for example, are considerably more migratory overall than those with only an elementary education, and nearly four times as likely to migrate across a state line. In general, each successively higher increment of education between these extremes is associated with a higher rate of migration.<sup>10</sup>

**Past migration experience.** Migration is frequently a repetitive episode, and observed mobility rates tend to reflect repeated and frequent moves by the same people rather than single moves by others. Consequently, people with a history of past moves show a disposition to move again.<sup>11</sup> There are several possible explanations of this hypermobility:

Some of the characteristics which make people likely to move are characteristics which persist. The same factors may continue to operate and cause people to move again. The first move may itself predispose people to the second. A few moves, about 8 percent, are regarded as failures by the people who move, and they may either return to their first home or try again somewhere else. Successful moves may lead to attempts to achieve further success by moving again. Not all moves are made with the expectation of permanence. A man may be transferred to a location with the knowledge that later he will move again. People who have moved are very likely to have left behind relatives and friends to whom they will report on conditions in the labor market in the new area which they left, which may increase the likelihood of additional out-migration.<sup>12</sup>

The overall effect of repeated movement can be quite extensive when a locality is undergoing rapid migratory growth. In San Jose, California, for example, the population has been built up by waves of immigration. (Migration furnished fully two-thirds of its growth during the 1960s.) As a result, San

Geographic Mobility among Professional Workers: A Multivariate Analysis," *Demography* 4, no. 1 (1967): 293-309; James Tarver, "Occupational Migrational Differentials," *Social Forces* 43 (December 1964): 231-241.

<sup>9</sup>Arlene S. Holen, "Effects of Professional Licensing Arrangements on Interstate Labor Mobility and Resource Allocation," *Journal of Political Economy* 73 (October 1965): 492-498.

<sup>10</sup>For comparative data, see U.S. Bureau of the Census, *Current Population Reports*.

<sup>11</sup>Sidney Goldstein, "The Extent of Repeated Migration: An Analysis Based on the Danish Population Register," *Journal of the American Statistical Association* 59 (1964): 1121-1132; Peter A. Morrison, "Chronic Movers and the Future Redistribution of Population," *Demography* 8 (May 1971): 171-184.

<sup>12</sup>John B. Lansing and William Ladd, *The Propensity to Move* (U.S. Department of Commerce, Area Redevelopment Administration, July 1964), p. 16.

Jose's population is heavily weighted with chronically mobile individuals and is thus subject to extraordinarily high rates of subsequent outmigration.<sup>13</sup>

### Why People Move

The migration differentials mentioned above show the broad sociological and economic influences on decision-making, but they reveal nothing directly about mobility intentions and decisions per se (that is, apart from their observed outcomes). To study directly the *intent* to move, we must examine would-be movers—people who fail to realize their preferences or intentions during the period of observation—as well as those who actually move.

In their landmark national survey, *The Geographic Mobility of Labor*,<sup>14</sup> Lansing and Mueller advance our insights into these behavioral aspects of mobility. This study has two valuable features: first, it examines the intent to move; second, it identifies the specific factors that motivate migration by different classes of workers and in economically different regions of the United States.

On the first point, the survey indicates that there is a sizable reservoir of would-be migrants—people who say they want to move but fail to realize their preferences or intentions. Lansing and Mueller estimate that for every actual migrant, nearly two people *expect* to migrate and four say they would *like* to migrate. In short, many people seem unable to bridge the gap between their inclination to move and action.

The survey also advances our understanding of people's reasons for moving. Unlike past studies, Lansing and Mueller's allowed for the possibility that people may move for more than one reason.<sup>15</sup> The findings indicate that people migrate fundamentally (but by no means exclusively) for economic and job-related reasons. "Three moves out of four were, according to people's explanations, at least partly economic in purpose, while one out of four was made for noneconomic reasons exclusively."<sup>16</sup>

Economic reasons for moving are especially characteristic of the young, the college-educated, and the professional, but among these economically purposeful migrants, particular reasons vary. Table 1 shows that blue-collar workers tend to move to avoid unemployment, to find work, or to obtain a steadier job; professional workers, on the other hand, move chiefly to increase

their remuneration and are more likely to be subject to subsequent outmigration.

The most common reasons for migration are the more purposeful migration, the more likely to be subject to subsequent outmigration.

Reasons for migration are the more purposeful migration, the more likely to be subject to subsequent outmigration. Middle-aged and older adults often migrate for economic reasons, while young adults migrate for social reasons.

In summary, the main conclusions are:

First, judgment is required in interpreting the results. Second, the groups of people who migrate are the well-educated, the young, and the professional, which are the groups most likely to move. The earnings of these groups are higher than those of the blue-collar workers.

Unemployment is a major reason for migration. The unemployed are more likely to move than the employed. The unemployed are more likely to move than the employed.

The crux of the problem is the face of the migration. The migration may reflect the face of the migration. The migration may reflect the face of the migration. The migration may reflect the face of the migration.

<sup>13</sup>Annual net migration into metropolitan San Jose averaged nearly 4 percent during the 1960s. Social Security Continuous Work History Sample data suggest that this net flow was composed of about twenty-one arrivals and seventeen departures each year per hundred residents (or nearly ten actual moves for each "net migrant" added). About seven of these seventeen outmigrants, though, were people who had moved into San Jose only the year before. Indeed, fully one-third of the migrants attracted to San Jose had moved away a year later.

<sup>14</sup>John B. Lansing and Eva Mueller, *The Geographic Mobility of Labor* (Ann Arbor: Survey Research Center, Institute for Social Research, 1967).

<sup>15</sup>Stated reasons for moving elicited ex post facto may contain an element of rationalization; still, they furnish direct information on why and how moving decisions are reached. This direct information is invaluable in sorting out the clues embedded in aggregate data on differential migration.

<sup>16</sup>Lansing and Mueller, p. 39.

<sup>17</sup>Ibid., p. 66.

<sup>18</sup>Ibid., pp. 68.



their remuneration. Both professional and other white-collar workers are more likely than blue-collar workers to migrate as a result of transfer or reassignment.

The more educated a migrant, the more likely he is to be economically purposeful. Nearly a third of those with some college education are transferred or reassigned.

Reasons for moving vary also with age, but in a complex way. Young adults often move to avoid unemployment or to find work—part of the initial process of searching and experimentation in early working years. Middle-aged people generally move for income-related reasons. With advancing age, migrants show a reduced interest in improving earnings; by age sixty-five, economic or occupational motives for moving are infrequent.

In summarizing these points, Lansing and Mueller draw two general conclusions:

First, judging by people's own explanations, the decision to move among members of the labor force is strongly dominated by job-related economic reasons. Secondly, economic incentives seem to play the greatest role among the groups in the labor force which have the strongest economic position—the well-educated, the middle-aged, and the white-collar workers. The groups which are in a weaker position may move because of serious economic pressures (such as lack of work), or occasionally because the employer initiates a move. The more optional types of moves, directed primarily toward higher earnings or professional advancement, are relatively infrequent among blue-collar workers, older people, and the less educated.<sup>17</sup>

**Unemployment.** An unemployed worker is more inclined to migrate than an employed one; but overall, the unassisted migratory response to unemployment proves to be weak and uneven.<sup>18</sup> It is weak because most workers with unemployment experience, whatever their inclinations, nevertheless do not move. And it is uneven because workers more prone to unemployment—those in blue-collar occupations, those with low skill and educational levels, and those of advanced age—tend to be more immobile than others.

The crux of the problem is that most people are reluctant to move, even in the face of prolonged and severe unemployment hardship. This reluctance may reflect a carefully considered personal judgment that values familiar surroundings above the prospect of employment elsewhere. But the weight of evidence suggests that what really prevents the unemployed worker from migrating is the *unawareness* or *absence* of choice. Residents of federally defined depressed areas express surprisingly little interest in moving away, despite the economic "push" of low incomes and high unemployment. This lack of interest must stem at least partially from people's erroneous impressions about relative economic conditions elsewhere. Lansing and Mueller's

<sup>17</sup>Ibid., p. 66.

<sup>18</sup>Ibid., pp. 68–72.



Table 1. Economic Reasons for Moving (Percentage of Heads of Families Who Moved in the Last Five Years\*)

| Occupation              | Transfer;<br>Reassignment | Unemployment;<br>New, More, or<br>Steadier Work | Higher Rate<br>of Pay;<br>Better Job | Other<br>Economic<br>Reasons | No Economic or<br>Occupational<br>Reasons Given | Total | Number<br>of<br>Cases |
|-------------------------|---------------------------|---|--------------------------------------|------------------------------|---|-------|-----------------------|
| Professional, Technical | 23%                       | 7%  | 50%                                  | 4%                           | 16%   | 100%  | 108                   |
| Other white-collar      | 41%                       | 14%   | 27%                                  | 4%                           | 14%   | 100%  | 102                   |
| Blue-collar             | 13%                       | 26%   | 25%                                  | 10%                          | 26%   | 100%  | 179                   |
| Education               | 9%                        | 19%   | 20%                                  | 12%                          | 40%   | 100%  | 78                    |
| Eight grades or less    | 13%                       | 21%   | 24%                                  | 13%                          | 29%   | 100%  | 208                   |
| High school             | 32%                       | 7%  | 36%                                  | 7%                           | 18%   | 100%  | 189                   |
| College                 |                           |   |                                      |                              |   |       |                       |
| Age                     |                           |   |                                      |                              |   |       |                       |
| 18-24                   | 9%                        | 33%   | 26%                                  | 11%                          | 21%   | 100%  | 78                    |
| 25-34                   | 25%                       | 11%   | 37%                                  | 8%                           | 19%   | 100%  | 164                   |
| 35-44                   | 24%                       | 11%   | 31%                                  | 9%                           | 25%   | 100%  | 93                    |
| 45-54                   | 20%                       | 13%   | 30%                                  | 12%                          | 25%   | 100%  | 69                    |
| 55-64                   | 20%                       | 20%   | 8%                                   | 13%                          | 39%   | 100%  | 51                    |
| 65 and over             | 0%                        | 3%  | 3%                                   | 13%                          | 81%   | 100%  | 35                    |
| Total                   | 20%                       | 15%   | 29%                                  | 10%                          | 26%   | 100%  | 474                   |

Source: John B. Lansing and Eva Mueller, *The Geographic Mobility of Labor* (Ann Arbor: Survey Research Center, Institute for Social Research, 1967), tables 18 and 19.

\*Excluding moves in and out of armed forces and moves to and from college.

icates that over half the residents of depressed areas think that where they now live are just as good as, or better than, anywhere else. One-third believe that in other places the pay is *lower*, and the chance, in their line of work.<sup>19</sup>

Information and help, matters can work out differently. Several federal relocation assistance programs have demonstrated that many depressed people willing to relocate can be helped to find jobs and increase earnings elsewhere.<sup>20</sup> The success of these programs suggests that the voluntary response to unemployment could be strengthened by subsidies, with advantages redounding to workers and distressed areas.

**community ties.** The decision to move, both at the stage of planning and at the point of action, is closely tied to the presence or absence of family ties. Of the one in four who mention family reasons for migrating, one-third move to be reunited with family members. People who have moved and then come back (return migration); or family members who have moved and then may join migrants at their destinations (chain migration); or the marked tendency for family members or friends at a distance to encourage migration, since they are in a position to reduce psychological and informational barriers that discourage or block potential moves. Job information, help with moving and settling into a new community, and other tangible assistance reduce the inertia of those who might otherwise hesitate.

#### Where Decisions Are Made

Decisions are circumscribed by many personal considerations that create inertia and, once a decision to move has been made, often evidence a lack of deliberation.<sup>21</sup> As noted above, there is a sizable reservoir of potential migrants who either may be reluctant to make concrete moving plans or may postpone carrying them out after they have been made. Other studies of the decision to move have confirmed this important point: recent migrants are known to be oriented toward future mobility and better able to make a moving plan or choice compared with those who have not moved.

107.

These programs are reviewed in Charles K. Fairchild, *Worker Relocation: A Review of Department of Labor Mobility Demonstration Projects* (Final Report of a Manpower Administration, U.S. Department of Labor, by E. F. Shelley and J. C. Smith, April 1970); and in Martin Schnitzer, *Regional Unemployment and the Workers* (New York: Praeger, 1970).

It also shows that a significant fraction of high school students in lagging areas are away. Locational preference surveys identify some of the factors associated with the decision to leave and the kinds of destinations young people typically choose. See John R. Henderson, *Location Preferences, Migration, and Regional Growth* (New York: Praeger, 1970).

Mueller, pp. 135-144.

Data on these points, see Lansing and Mueller, table 70.

Van Aradol, Jr., et al., "Retrospective and Subsequent Metropolitan Residency," *Demography* 5 (1968): 249-267.

How fully do migrants deliberate in reaching a moving decision? How do they inform themselves about employment prospects where they are going? Many moves, it seems, are preceded by surprisingly little deliberation, and most potential movers consider only a narrow range of destinations and jobs:

One-third of migrants report that they began thinking seriously about moving only a month before they actually departed.

Two-thirds considered no other destinations than the places to which they actually moved.

Three-fifths relied on only one source of information to explore job opportunities in the new place.

Job information was most frequently obtained from friends and relatives (49 percent) or through special trips to look the situation over (33 percent). Migrants rarely consulted state employment agencies (6 percent).<sup>25</sup>

Thus, many migrants display a poor facsimile of rational choice behavior in reaching a decision to move. They decide hastily and give little or no thought to alternatives. Moreover, because they rely so heavily on family and friends in deciding where to go, migrants often go only to the places where friends and relatives have already settled.<sup>26</sup> The typical migrant's information sources, then, do not assure that he will move effectively.

#### Where Migrants Go

Migrants tend to move to large centers of population nearby. The relatively stable determinants of population size and distance fix the approximate gross volumes of migratory flows that connect any two places. Beyond these gravitylike effects, though, local economic conditions systematically influence where migrants go. Mobile people flow toward labor markets, especially nearby ones, that are attractive to job-seekers.<sup>27</sup> These points are examined in greater detail in the following sections, which examine migration's role in alleviating area distress and in shaping national metropolitan growth.

#### Migration from Distressed Areas

National economic growth and social change inevitably bring about mismatches of labor supply and demand. Job opportunities multiply in one region and disappear in another; new industries thrive while old ones wither. Migration acts as a bridge over which workers pass from areas where jobs are scarce to areas where workers are scarce. The following discussion centers

<sup>25</sup>Lansing and Mueller, p. 210.

<sup>26</sup>This is why we find enclaves of people, for example, from one section of Kentucky living in a particular neighborhood in Cincinnati. Such "beaten paths," which accord with historical movements of population, are perpetuated through the location of friends and relatives.

<sup>27</sup>See, for example, Michael J. Brennan, *Regional Labor and Capital Migration*, Final Report to the Area Redevelopment Administration, U.S. Department of Commerce, Contract No. C-325-65, May 1967, p. 73; Richard F. Muth, "Migration: Chicken or Egg?" *Southern Economic Journal* 37 (January 1971): 295-306; and Lowry, pp. 30-31.

on two as-  
balance, ar  
**Adjustment**

There is  
effective m  
this adjust  
stick are e  
that migra  
It is not cl  
conditions  
destination

Is migra  
more than  
there is litt  
policy tow

A cornerst  
outmigrati  
as well as o  
have been  
opportunit  
aim to pro  
asymmetry  
retaining tl  
outsiders in

The first  
origin was  
tion.<sup>28</sup> Nur  
conspicuous  
and insensi

This inte  
which illun  
probed peo  
or leaving.  
nondepress  
former gro  
more inclin

But othe

<sup>28</sup>William Al  
*ings of the Re*  
merce, Econo

<sup>29</sup>Lowry.

<sup>30</sup>William A  
Urban and  
California, A

<sup>31</sup>Lansing an

pects of migration: (1) how it operates to restore labor market and (2) the consequences of its selectivity.

#### **Aspects in the Labor Market**

There is a virtually unanimous agreement that labor force migration is an important mechanism of adjustment. Evidence conflicts, however, as to whether migration works symmetrically—that is, whether the carrot and the stick are equally important as motivating forces. All studies seem to agree that migrants find their way to destination areas where labor is in demand. They fear, however, whether their departure is stimulated by depressed conditions at origin. Although migrants go to the economically “correct” areas, they may not as readily leave from the “correct” origins.

Is migration’s wisdom two-sided or only one-sided? The question is of academic interest. The one-sided interpretation, suggesting that there is little or no economic “push,” contradicts the logic of existing national and local distressed areas:

One assumption . . . is that economic distress leads to heavy outmigration, and that this inflicts upon the area of origin grave social costs. Policies of investment in depressed areas based in large measure on an attempt to expand local economic activities and prevent economically-forced outmigration. In effect, they deprive people the freedom to stay in their own region. . . . [But if] development programs would not have the effect of displacing the original inhabitants, but rather that of increasing the inflow of people to areas which typically already suffer from a labor surplus.<sup>28</sup>

The indication that outflow rates may be unresponsive to “push” at origin is given in Lowry’s well-known study of intermetropolitan migration. Numerous other studies have reported similar findings that point to a weak or absent relationship: outmigration seems largely spontaneous and responsive to local labor market conditions.<sup>29</sup>

This interpretation is supported also by Lansing and Mueller’s survey, which analyzes the behavioral mechanism of this asymmetry. The survey explores people’s perceptions of origin conditions and their motives for staying or leaving.

Comparisons between residents of depressed areas and those of nondepressed areas revealed no obvious “push” for outmigration among the depressed.<sup>31</sup> As previously noted, people in depressed areas were no more likely to leave than their counterparts in nondepressed areas.

Other studies report finding the expected relationship between outmi-

Lonso, “Policy Implications of Intermetropolitan Migration Flows,” *Proceedings of the Regional Economic Development Research Conference* (U.S. Department of Commerce, Economic Development Administration, April 19, 1972), p. 146.

Lonso, “The System of Intermetropolitan Population Flows,” Institute of Regional Development, working paper no. 155 (Berkeley: University of California, August 1971).

and Mueller, table 106.

gration and economic conditions at origin. Their claim is that origin "push" surely is operative but that its role is masked by improper measures of true unemployment conditions. Average or end-period rates of unemployment at origin are poorly suited to explaining differential outmigration, for such rates have already been modified by the most recent corrective effects of outmigration. Synthetic measures like "prospective" or "potential" unemployment are needed to avoid this problem. They are designed to relate outmigration to the total pressure for outflow, rather than to the residual level of unemployment that remains after surplus workers depart.<sup>32</sup>

Evidence of economic "push" does appear in studies that rely on these synthetic measures. Olvey reports "a highly significant relation between outmigration rates and local labor market conditions, as characterized by 'prospective unemployment' . . . [that is,] the level of unemployment which would have accumulated by the end of the period had no outmigration taken place."<sup>33</sup> And Mazek reached a similar conclusion using a comparable synthetic measure ("potential unemployment").<sup>34</sup>

How can these conflicting findings about the authenticity of economic "push" be reconciled? Neither side of the argument is convincing, the one marred by an improperly specified unemployment variable, the other calling for a heroic reliance on a synthetic construct. Nevertheless, the issue has key policy implications; and until these processes can be fully sorted out, Alonso notes, "there is a good chance that our policies and programs may not only be wasteful or ineffective, but indeed counterproductive."<sup>35</sup>

Thus far, ironically, the evidence on both sides rests partly or wholly on five-year migration data from the 1960 census. The problem is that five-year migration flows, viewed in cross section, capture long-term structural differences among areas but disguise short-term adaptive changes.

Examined cross-sectionally, outmigration rates are equally high in prosperous metropolitan areas and depressed ones. But the high outflows from

<sup>32</sup>See Warren F. Mazek, "The Efficacy of Labor Migration with Special Emphasis on Depressed Areas," working paper CWR 2 (St. Louis: Washington University, June 1966), chap. 3; and Lee Donne Olvey, "Regional Growth and Inter-regional Migration: Their Pattern of Interaction" (Ph.D. diss., Harvard University, Department of Economics, 1970), p. 85. Both use variants of "prospective unemployment" as originally formulated by Cicely Blanco, "Prospective Unemployment and Interstate Population Movements," *Review of Economics and Statistics* 46, no. 2 (May 1964): 221-222.

<sup>33</sup>Olvey, p. 107.

<sup>34</sup>Olvey and Mazek use different operational definitions. Olvey defines prospective unemployment as "the increase in the labor force (proxied by net population increase) minus total employment increase, assuming a zero rate of outmigration. It is a measure of the increase in unemployment which would occur if outmigration dropped to zero while all the other variables held constant" (p. 85). Mazek defines "potential unemployment" as "the unemployment rate which would exist in a region at the end of the period studied if no migration—in or out—took place during the period. It is estimated by first estimating the size of the labor force at the end of the period, assuming no migration, and subtracting from this the actual employment of the labor force at the end of the period. . . . This is the precise equivalent of adding the net outmigration of the labor force that took place over the period to the actual unemployment rate at the end of the period" (p. 21).

<sup>35</sup>Alonso, "Policy Implications of Intermetropolitan Population Flows," p. 149.

fast-growing San Jose, Houston, Atlanta, and the like, may result simply from the hypermobile population base that has been built up by wave after wave of immigrants.<sup>36</sup> Conversely, the declining Johnstowns and Pittsburghs have already lost many of their mobile residents; therefore, their rates of outflow would be only nominal, even though by economic reasoning these rates should be higher than elsewhere. Consequently, if "push" does operate in the short term, any trace of its effect may be obscured by the opposing structural effect of hypermobility.

Better data are needed so that more pointed questions can be asked: Do *annual* outflows respond to *annual* fluctuations in economic conditions? Is migration reacting to this year's growth, accommodating to last year's growth, or anticipating next year's growth? Where approximations to such data have been used, the results have been revealing. Metropolitan outmigration, although economically insensitive over the long term (as Lowry found), seems responsive to short-run changes in local employment growth.<sup>37</sup>

#### **Consequences of Selective Migration**

Outmigration acts as an economic adjustment mechanism by reducing local labor surpluses and lessening competition for scarce employment. But what begins as an economic adjustment can become over time a gradual leaching away of local human capital. At some point, outmigration accelerates local economic distress by reducing productivity and, hence, attractiveness to new industry.

Since outmigration usually draws away more highly qualified members of the labor force—the young, the educated, and the skilled—the labor force left behind tends to be overaged, undereducated, and underskilled. This effect often is further accentuated by immigration of people similar to those who have remained behind.<sup>38</sup> As a labor force declines in quality, distressed areas become less attractive to new industries that require a supply of skilled workers. Only marginal firms paying low wages want an overaged, undereducated, and underskilled labor force. Where downside rigidity has kept wages high relative to productivity, an area fails to attract new employers and hence continues to lose labor, though perhaps too slowly.<sup>39</sup>

Furthermore, since the people who stay are generally the less migration-prone, the remaining population shows a gradually reduced potential for mobility. This means that stronger and stronger economic incentives would be necessary to induce additional people to move away in order to maintain any balance between population and shrinking employment.

Prolonged and heavy outmigration, then, leaves behind those who are least able to cope with the unfavorable conditions that led others to depart

<sup>36</sup>Morrison, "Chronic Movers."

<sup>37</sup>For evidence on this point, see Vernon Renshaw, "The Role of Migration in Labor Market Adjustment" (Ph.D. diss., Massachusetts Institute of Technology, Department of Economics, 1970).

<sup>38</sup>Lansing and Mueller, pp. 318–319.

<sup>39</sup>Olvey, pp. 127–129.



Table 2. Characteristics of Population in Redevelopment and Nonredevelopment Areas

|                                       | Redevelopment Area |            | Nonredevelopment |
|---------------------------------------|--------------------|------------|------------------|
|                                       | High-Unemployment  | Low-Income |                  |
| Age                                   |                    |            |                  |
| Under 35                              | 19%                | 15%        | 26%              |
| 35-54                                 | 42                 | 39         | 41               |
| 55 and over                           | 39                 | 46         | 33               |
| Total                                 | 100                | 100        | 100              |
| Education                             |                    |            |                  |
| Eight grades or less                  | 39                 | 54         | 27               |
| High School                           | 47                 | 32         | 46               |
| College                               | 14                 | 14         | 27               |
| Total                                 | 100                | 100        | 100              |
| Labor force status                    |                    |            |                  |
| Wage and salary workers               | 60                 | 43         | 65               |
| Self-employed and farmers             | 11                 | 24         | 11               |
| Not in the labor force                | 29                 | 33         | 24               |
| Total                                 | 100                | 100        | 100              |
| Occupation of wage and salary workers |                    |            |                  |
| Professional, technical               | 11                 | 16         | 19               |
| Managerial, officials                 | 6                  | 7          | 10               |
| Clerical and sales                    | 13                 | 9          | 17               |
| Craftsmen and foremen                 | 26                 | 16         | 20               |
| Operatives                            | 25                 | 25         | 20               |
| Laborers and service                  | 19                 | 27         | 14               |
| Total                                 | 100                | 100        | 100              |

Source: Lansing and Mueller, *Geographic Mobility of Labor*, pp. 303, 305, 307.

in the first place. The remaining residents tend to lack the attributes and skills that would attract new employers who could offer them jobs or that would predispose them to move away as others before them did.

The data from Lansing and Mueller's study show the effects of selective outmigration. Using a supplemental sample of interviews with residents of redevelopment areas (classified by the Area Redevelopment Administration as "depressed"), the authors compared families in high-unemployment and low-income areas with families outside these areas.<sup>40</sup> Their findings, shown in the data of table 2, are as follows:

Age. Redevelopment areas contained proportionately fewer people under thirty-five and more fifty-five and over. Low-income areas contained an exceptionally large fraction (46 percent) of people fifty-five and older.

Education. The proportion of college-educated family heads was only half as large in the redevelopment areas. In low-income areas, 54 percent had only grade school education, compared with 27 percent in nonredevelopment areas.

<sup>40</sup>Lansing and Mueller, p. 292. Areas that are characterized by high and persistent unemployment are known as 5A; areas that have a high percentage of low-income families are known as 5B, and are predominantly rural counties or small labor market areas. At the time of the survey and in the preceding four years, both 5A and 5B areas suffered relatively severe economic distress.

Labor force areas, fewer were under

On the ir typical cross migrants w self-employ outmigrants tendency fo centuated.<sup>41</sup>

A final p gration from ished produ Yet localiti significant f ratio of job workers, ir filled by ex elsewhere, comparative

#### Migration a

Historica attention to increasingly the bulk of growing cen growth, and migrants at

Since the t metropolita metropolita substantiall recently, in a result of number of number of r

<sup>41</sup>Lansing and  
<sup>42</sup>For a specif Rural Area," migration effe for Regional I  
<sup>43</sup>William Ah Century Ame ed., Niles Ha

Labor force status and occupation. Compared with nonredevelopment areas, fewer family heads were in the labor force, and white-collar workers were underrepresented.

On the inflow side, people migrating to redevelopment areas were not a typical cross section of all migrants. Redevelopment areas tend to draw migrants who are older, undereducated, out of the labor force, blue-collar, self-employed, or farmers. Thus, as immigration replaces the more productive outmigrants with people who may be less productive economically, the tendency for redevelopment areas to lose valuable human capital is accentuated.<sup>41</sup>

A final point concerns the effects of return migration. Prolonged outmigration from labor surplus areas, while leaving behind a labor pool of diminished productive potential, does reduce competition for scarce employment. Yet localities from which outmigration has been extensive are subject to significant flows of return migrants which may largely cancel the increased ratio of jobs to workers. Jobs vacated by outmigrants may pass on to local workers, improving their welfare, but it may be that their jobs are usually filled by exresidents who, having gained industrial skills and experience elsewhere, return home and are hired in preference to the resident (and comparatively inferior) labor force.<sup>42</sup>

#### Migration and National Metropolitan Growth

Historical analysis of twentieth-century American urbanization has called attention to an interesting emerging pattern: the national migration system increasingly favors relatively few metropolitan centers that together receive the bulk of available migratory growth. More specifically, the most rapidly growing centers have accounted for an increasing share of total metropolitan growth, and this increase is most pronounced for centers that have drawn migrants at unusually high rates.

Since the beginning of the century . . . a very large share of American metropolitan growth, and a far larger share of the net immigration into metropolitan areas, have been absorbed by those metropolises that grew substantially faster than the metropolitan set. This share has been increasing recently, in spite of the declining importance of metropolitan immigration, as a result of a more active and selective intermetropolitan migration. As the number of areas with substantial net immigration has increased, so has the number of metropolises which are net exporters of people.<sup>43</sup>

<sup>41</sup>Lansing and Mueller, p. 319.

<sup>42</sup>For a specific case study, see Irwin Gray, "Employment Effect of a New Industry in a Rural Area," *Monthly Labor Review* 92 (June 1969): 26-30. A fuller discussion of the return migration effect is given in Niles M. Hansen, *Rural Poverty and the Urban Crisis: A Strategy for Regional Development* (Bloomington: Indiana University Press, 1970), pp. 262-263.

<sup>43</sup>William Alonso and Elliott Medrich, "Spontaneous Growth Centers in Twentieth-Century American Urbanization," in *Growth Centers in Regional Economic Development*, ed., Niles Hansen, (New York: Free Press, 1972), pp. 247-248.



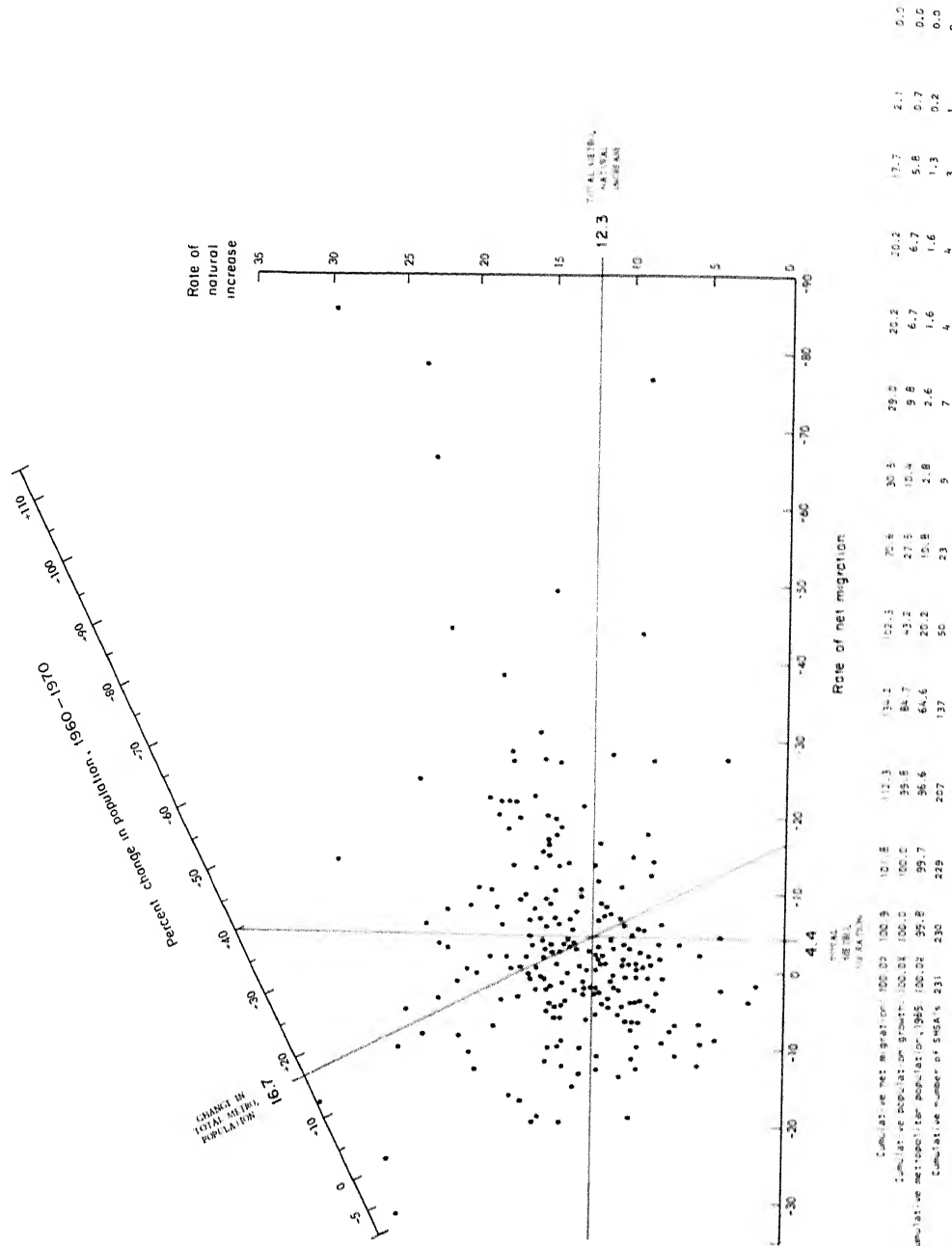


Figure 2

Building on selective and continued during

Between 1 Statistical A is the combination (a

"Spontaneous areas that a of a "spontaneous is, with designed to of available 1970 decade growth, head SGCs to w residents per percent of the percent of the decade growth. The residents per 102.3 percent positive rate siphoned out alike.) Although, these fit Thus, much net immigration centers continue.

In the pastional migration, along with centers somewhat reduced only in these traditions evidently become or decline.<sup>45</sup>

The dynamic policies might effect of migration

<sup>44</sup>All SMSAs, Economic Areas

<sup>45</sup>William Alonso

ilding on Alonso and Medrich's earlier work, my own analysis shows this active and uneven pattern of national metropolitan growth to have continued during the last decade.

Between 1960 and 1970, the population of all 231 Standard Metropolitan Statistical Areas (SMSAs) increased by 16.7 percent. (see fig. 2.) This growth is the combined result of natural increase (averaging 12.3 percent) and net migration (averaging 4.4 percent).<sup>44</sup>

"Spontaneous growth centers" (SGCs) are defined here as metropolitan areas that are drawing migrants at unusually high (net) rates. The concept of a "spontaneously" growing center is flexible; hence, its operational definition is, within limits, arbitrary. The cumulative totals beneath figure 2 are designed to show, for alternative SGC definitions, the disproportionate share of available migratory growth acquired by certain SMSAs during the 1960-1970 decade. These data reveal a markedly uneven distribution of migratory growth, heavily favoring SGCs. Consider, for example, the twenty-three SMSAs to which net immigration, 1960-1970, added twenty or more new residents per hundred population in 1960 (fig. 3). They contained only 10.8 percent of the entire metropolitan population in 1965, yet they drew 70.6 percent of the cumulative net migration that fed metropolitan growth during the decade and accounted for 27.5 percent of all metropolitan population growth. The fifty SGCs to which net immigration added ten or more new residents per hundred population captured a cumulative net migration of 43.3 percent. (Since the lowest net migration rates are negative, the sum of positive rates exceeds 100 percent; this merely indicates that migrants were drawn out of nonmetropolitan areas and non-SGC metropolitan areas as well.) Although representing only 20.2 percent of the metropolitan population, these fifty SGCs account for 43.2 percent of all metropolitan growth. Thus, much of the growth in U.S. metropolitan areas, and a large portion of net immigration feeding that growth, is confined to a few rapidly growing centers containing only a comparatively small part of the nation's population.

In the past, urban centers grew mainly through rural-to-urban and international migration. These large migrations from outside the metropolitan system, along with the substantial cushion of natural increase, afforded all urban centers some measure of growth; migration *among* metropolitan areas provided only marginal rate differences. But now, with the cushion provided by these traditional forces shrinking rapidly, intermetropolitan flows have increasingly become more influential determinants of local population growth or decline.<sup>45</sup>

The dynamics of migration have an important bearing on how public policies might influence the redistribution of population. Ordinarily, the net effect of migration is small in relation to the much larger (but, for the most

<sup>44</sup> SMSAs are as of 1970; those in New England are defined as Metropolitan Statistical Areas.

<sup>45</sup> William Alonso, "Policy Implications of Intermetropolitan Migration Flows," p. 143.

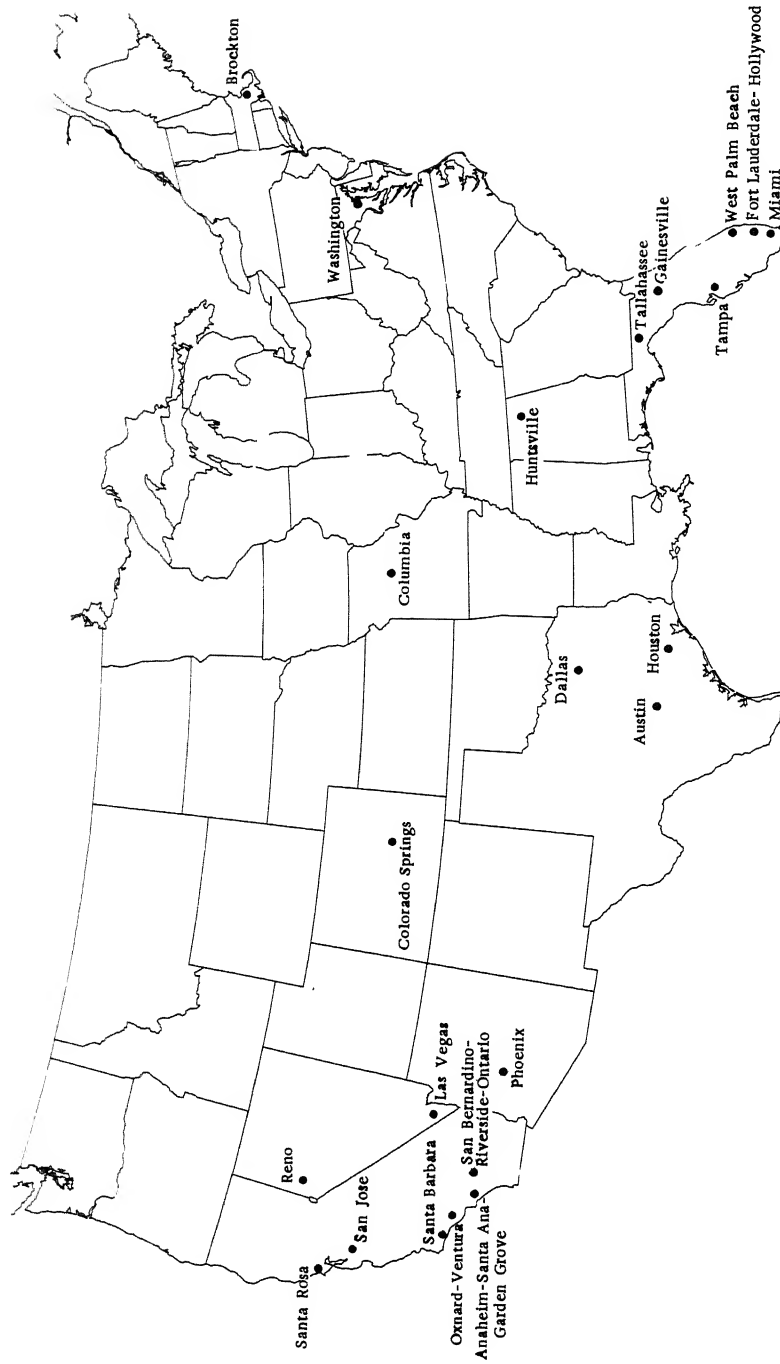


Figure 3. Locations of twenty-three Top-ranking Spontaneous Growth Centers

part, self-balancing) gross flows into and out of a locality. But the flow into an area—the “pull” of economic attraction—is potentially responsive to policy measures. In theory, population redistribution could be influenced by policies designed to enhance migrants’ preferences for certain destinations over others.

Public policies already seem to be arbitrarily influencing the economic attractiveness of certain areas. Agencies that build highways, award defense contracts, and site federal installations are simultaneously redistributing employment growth and altering incentives for private investment. These and other programs and activities, although nominally unrelated, have in reality accumulated and exert a powerful but undirected migratory influence. The “spontaneous” growth centers demonstrate this fact.

Defense expenditures, for example, have been a major determinant of the location of economic (and hence migratory) growth. The geographic pattern of these expenditures clearly favors the nation’s fastest-growing counties (table 3). Defense payrolls, totaling \$20.8 billion in fiscal 1969, were heavily concentrated in the fastest-growing counties and underrepresented in the slowest-growing ones; and defense contracts, totaling an additional \$38.8 billion, favored the fastest-growing counties and those that were richest and

Table 3. Concentration of Defense Payrolls and Defense Contracts, Fiscal 1969

| Type of County               | (1) % of U.S.<br>Population | (2) % of U.S.<br>Program | (2) ÷ (1)<br>Concentration<br>Ratio |
|------------------------------|-----------------------------|--------------------------|-------------------------------------|
| Defense Payrolls             |                             |                          |                                     |
| Poorest <sup>a</sup>         | 10.0                        | 1.8                      | 0.18                                |
| Richest <sup>a</sup>         | 10.1                        | 6.9                      | 0.68                                |
| Slowest-growing <sup>b</sup> | 10.2                        | 4.4                      | 0.43                                |
| Fastest-growing <sup>b</sup> | 10.0                        | 22.4                     | 2.24                                |
| SMSAs > 1 million in 1966    | 37.3                        | 29.6                     | 0.79                                |
| SMSAs < 1 million in 1966    | 29.6                        | 44.3                     | 1.50                                |
| Non-SMSA urban               | 11.0                        | 10.3                     | 0.94                                |
| Rural                        | 22.1                        | 15.8                     | 0.71                                |
| Defense Contracts            |                             |                          |                                     |
| Poorest <sup>a</sup>         | 10.0                        | 1.5                      | 0.15                                |
| Richest <sup>a</sup>         | 10.1                        | 13.5                     | 1.34                                |
| Slowest-growing <sup>b</sup> | 10.2                        | 7.6                      | 0.75                                |
| Fastest-growing <sup>b</sup> | 10.0                        | 16.2                     | 1.62                                |
| SMSAs > 1 million in 1966    | 37.3                        | 50.9                     | 1.36                                |
| SMSAs < 1 million in 1966    | 29.6                        | 35.4                     | 1.20                                |
| Non-SMSA urban               | 11.0                        | 7.4                      | 0.67                                |
| Rural                        | 22.1                        | 6.3                      | 0.29                                |

Source: U.S. Office of Management and Budget, Evaluation Division, “Locational Analysis of Federal Expenditures in Fiscal Year 1969,” Report of the Joint OMB-EDA Locational Analysis Project, September 1, 1970 (mimeograph), table 2.

<sup>a</sup>Based on 1966 per capita income.

<sup>b</sup>Based on population growth, 1960–1966.

metropolitan.<sup>46</sup> Since defense payrolls and contracts amounted to fully 30 percent of total reported federal expenditures in fiscal 1969, these disparities probably had a strong effect in initiating economic and migratory growth in certain favored localities. Indeed, a study of the relationship between Department of Defense contracts and population growth in California noted a direct correlation between the two on a county-by-county basis.<sup>47</sup>

Defense spending must exert a sizable influence on the location of growth and, as such, contribute to a *de facto* policy of selective growth. Both the magnitude of expenditures and their concentration in the fast-growing counties make them a cornerstone of the present "spontaneous" growth center pattern.

### Conclusion

Migration produces many complex and subtle changes that are felt most immediately at the local level but that are ultimately felt nationally as well. In recent years, three general trends in the population's movement have attracted attention because they appear to be closely related to the emergence of economic, environmental, and social problems. The discussion has reached a point where, it is said, improving amenities in cities, achieving equity in race relations, and preserving the integrity of rural areas all depend on instituting national policies to guide population redistribution.

The first trend is the concentration of people in a few large metropolitan regions, which evokes concern over congestion and environmental deterioration. The second is the emerging pattern of racial segregation along city-suburban boundaries, which seems to intensify the economic burdens of the city and confound efforts at integration. The third is the migration away from economically distressed rural areas, which leaves behind those people—the undereducated and the elderly—who are least qualified to cope in any economic environment. This last trend has already inspired policy measures, but these have had little success so far in restoring rural economies. Together, these three trends reflect a basic demographic transformation that has been under way for several decades as a result of migration. But because each trend is associated with things gone wrong, the geographical settlement patterns themselves have come to be labeled "problems."

Some observers say the solution to these problems lies in changing the geometry of overall settlement patterns or at least in reducing the flow of migrants into large metropolitan centers. Others claim that it is unfeasible, and probably unwise, to interfere with the population's distribution. They argue that its spatial arrangement and rearrangement represents the ter-

<sup>46</sup>U.S. Office of Management and Budget, Evaluation Division, "Locational Analysis of Federal Expenditures in Fiscal Year 1969," Report of the Joint OMB-EDA Locational Analysis Project, September 1, 1970 (mimeograph).

<sup>47</sup>James L. Clayton, "Defense Spending: Key to California's Growth," *Western Political Quarterly* 15 (1962): 280-293.

ritorial counterpart of a tightly meshed national metropolitan economy. Since population shifts are a slow process, responsive to the workings of this economic system, achieving significant changes in distribution would require decades of sustained intervention. Furthermore, the goals of such intervention are now only poorly articulated and perhaps even misinformed.

My own view is that there are realistic and effective ways to guide the migration system as a whole toward achieving desirable ends, and these emerge from a better understanding of how migration works: what causes it to occur; what its effects on migrants are; and how it affects the places they come from and the places to which they go.

#### **Technical Appendix: Weaknesses in Operational Definitions and Data Collection Practices**

Studies of U.S. migration suffer from common weaknesses in operational definitions and data collection practices. Since these weaknesses often affect research conclusions, they are summarized briefly here.

##### **Analytical Definitions**

A migrant is customarily defined as a person who moves *during* a fixed time interval, perhaps one year or five years; a nonmigrant is a person who does not. But this operational definition has an analytical weakness: it includes as nonmigrants both the genuinely immobile person who has, in effect, "pledged allegiance" to his present community, and the potentially mobile person who simply has not *yet* migrated. This is like defining the labor force to include only those who are now employed. The standard definition is more discriminating: the labor force consists of all people who are employed *or seeking work* at the given time. Likewise, the migrant class should consist of all people who actually move *or seek to move*; and corresponding to the class of "unemployed" people, we might speak of a class of "unmigrated" people.

##### **Classification of People**

Migration studies are concerned chiefly with the factors that predispose or guide movement and the consequences that ensue. Ideally, the information pertaining to a move should be linked to the event itself, as is done for births and deaths. Population registers in some European countries (such as the Netherlands) use this type of reporting, but virtually all accounting of population movements in the United States is on a calendar basis. That is, a person's current place of residence is compared with his earlier places of residence at arbitrary points in time (a year ago, five years ago, at birth, and so forth). We thereby know only where he resided on two or more particular dates, and we then make inferences about his actual moves. Because we infer moves, not count them directly, our measures cannot register every move that occurs.

For example, data from the five-year migration question asked on the censuses of 1960 and 1970 leave major uncertainties about actual patterns and frequencies of movement. Quite diverse patterns of migration are masked by inadequate statistics: people who moved once are lumped in with

those who moved repeatedly; and those who migrated and then returned are included with those who did not move at all. Thus, the conclusions drawn from these data are asymmetric and biased: where census data indicate a move, movement *has* occurred; where they fail to indicate a move, movement still *may* have occurred.

Census data cannot indicate the *number* of moves, but only the fact that there was at least one. For example, the 1960 census question on interstate migration status, 1955-1960, differentiates only the following two types:

1. Migrated not at all; or at least twice, returning to origin (1955 and 1960 states of residence agree).

2. Migrated at least once (1955-1960 states of residence disagree).

When place of birth is added, the possibilities expand to three:

1. May have migrated, but lived at birthplace in 1955 and in 1960 (birthplace = 1955 residence = 1960 residence).

2. Migrated at least once (birthplace = 1955  $\neq$  1960; or birthplace  $\neq$  1955 = 1960).

3. Migrated at least twice (three different states named).

Types 1 mix nonmigrants and return migrants; types 2 mix the one-time mover with the footloose chronic mover.

The analytical precision of migration research has increased. With the recent availability of survey data and residence histories, such information has made it possible both to record moves directly and to probe decision-making processes. With new insight into the socioeconomic determinants of the *intent* to move, it became possible to trace how a potential migrant develops into an actual migrant and to identify factors that prompt and impede movement. Residence histories illuminated the sequences of moves more directly than before, so that single moves could be examined within the context of a series of related acts.

#### Spatial Classifications

Since migration must be defined with reference to some geographic boundary, its operational specification depends directly on the type of boundaries chosen. Demographers distinguish two types of mobility: (1) local (or residential) mobility, referring to moves within the same community, and (2) migration, referring to moves between communities. Clearly, the precision of this classification hinges on the operational definition of "community."

In current practice, the Bureau of the Census distinguishes the population classes shown in the table below according to mobility status and type of mobility. Classifying anyone who moves across a county line as a migrant has one unavoidable drawback: because metropolitan areas consist of clusters of integrated counties, many intercounty moves are simply changes in residence within the same larger "community."

To avoid this problem, some analysts prefer to define migration as moves across the boundaries of labor market areas (as delineated by the U.S. Department of Labor). Since these boundaries generally coincide with those of metropolitan areas, this definition has the advantage of excluding in-

#### Mobility Status

Living in same  
Living in different  
Living in different  
Same state  
Different state  
Contiguous  
Noncontiguous

transmetropolitan  
migration

Finally,  
miles or more

#### The Statistics

For a large  
departure:  
community  
tion is a series  
or departure.

General  
powerful  
casually  
and going  
California  
people move  
net migration  
Pittsburgh  
of the major  
treat network  
menon of  
behavior  
other people

| <i>Mobility Status</i>                   | <i>Type of Mobility</i>     |
|--|-----------------------------|
| Living in same house                     | Nonmover                    |
| Living in different house in same county | Intracounty mover           |
| Living in different county               | Migrant                     |
| Same state                               | Intrastate migrant          |
| Different state                          | Interstate migrant          |
| Contiguous states                        | Contiguous state migrant    |
| Noncontiguous states                     | Noncontiguous state migrant |

trametropolitan moves across county boundaries from the definition of migration.

Finally, migration has been defined in certain studies as moves of fifty miles or more.

#### **The Statistical Fiction of "Net Migration"**

For a locality, the arithmetic difference between migratory arrivals and departures is an important change-in-stock factor and a crude proxy for the community's "attractiveness." For analytical purposes, though, net migration is a statistical fiction. No person is a "net migrant"; he can only arrive or depart.

Generally, net migratory gains or losses are only the surface ripples of powerful crosscurrents modifying a locality's labor force. Indeed, what we casually regard as a "community" is actually a procession of people coming and going. Between 1965 and 1970, for example, metropolitan San Jose, California, gained 75,000 new residents through migration, but 395,000 people moved either in or out. Metropolitan Pittsburgh, Pennsylvania, had a net migration loss of 72,000 residents—still, 138,000 persons moved *to* Pittsburgh. Consequently, the one-way drift of net migration is no measure of the many migratory moves that people undertake, and it is erroneous to treat net migration—such as Pittsburgh's 72,000 net loss—as the phenomenon of people's behavior to be explained. Rather, we must explain the behavior of 138,000 people who moved to Pittsburgh and that of 210,000 other people who moved away.



# 10 Tanzania 1920-63: The Spatial Impress of the Modernization Process\*

Peter R. Gould

"From the combination of all these data, we might eventually get a solid structural and quantitative basis. . ."

Karl Deutsch, *Nationalism and Social Communication*.

"A fundamental intellectual implication of a world of change is the greater theoretical utility of the concept of process over that of structure in sociological and cultural analysis."

Emmanuel Mesthene, "How Technology Will Shape the Future," *Science*, No. 3837, 1968.

"Once the processes are understood, the structures manifested at given time-points will emerge with even greater clarity."

Evon Vogt, "On the Concepts of Structure and Process in Cultural Anthropology," *American Anthropologist*, 1960.

THE charge to examine old problems with new eyes, and so obtain a firmer understanding of the *structure* of the modernization process, came from Karl Deutsch in 1953. And now, sixteen years later, other social and behavioral scientists urge us to focus upon the *process*, the very dynamics of modernization, noting that with this more powerful view of change the sequential structures themselves will be clarified. But in examining the modernization process in Africa, we feel almost helpless in trying to reply to such a call—a call that essentially requires us to link Space and Time and Man in such a way that a general system is created to postdict the course of a modernizing society. It is difficult to see how we can write good theory illuminating the *process* of modernization before we have good descriptions of the basic spatial patterns that ultimately must be linked together through time. No social scientist, no matter what disciplinary viewpoint he assumes, will feel that his work is being disparaged when I say that our knowledge of

\* Revised from a paper delivered at the annual meeting of the American Political Science Association, Washington, D.C., September 1968. Copyright 1968, The American Political Science Association.

For supporting field work in Tanzania, I should like to acknowledge a Cross-Cultural Research Grant from Syracuse University, while library work in London and computing time was supported by The Pennsylvania State University. Grateful acknowledgment is made to Dr. Edward Soja for providing me with flow data on Tanzania, and to Mr. Thomas Leinbach for assisting in all the computer operations.

Reprinted from *World Politics*, vol. 22, no. 2 (January 1970), pp. 149-170, with permission of the author and Princeton University Press. ©1968 by Princeton University Press.

T  
such seq  
process :  
of a gec  
terizing  
points i  
sequence  
of chang

After  
nineteen  
veyed n  
known  
343,000  
eter, un  
later, in  
colonial

The c  
degree c  
was a ce  
ment in  
slaughte  
litical co  
to sixty-  
of admi  
Europe:  
nineteen  
by rail i  
was ope  
from 18  
Rail lin  
built in  
campaign  
British  
complet  
and cor  
Thus th

<sup>1</sup> Carl  
1904), 3.

quential patterns in Africa is slight, while our understanding of it is virtually nonexistent. Thus, from the spatially biased view of the geographer, I seek a sequence of measured descriptions characterizing the structure of an impressed, modernizing system at five points in time over the space called Tanzania. Such a descriptive sequence constitutes the first step on the road to modelling the *process* of change in Africa.

#### DELIMITATION AND DEVELOPMENT TO 1919

As a result of a series of diplomatic and military maneuvers typical of the late nineteenth century in Africa, blue lines were drawn across poorly surveyed maps in the chancelleries of Europe to delimit the area now known as Tanzania. Such cursory delimitation brought an area of 363,000 square miles, nearly a thousand miles across its greatest diameter, under the control of a single colonial regime. Only seventy years later, in 1961, the independent nation of Tanzania was born from the ruins of German and mandate rule.

The early years of such control are difficult to trace with a high degree of spatial resolution. From the turn of the century, German rule was a contradiction of enlightened economic and educational development in a few areas, with barbaric, punitive expeditions in others that killed tens of thousands and laid waste whole countrysides. Political control demanded a system of communications, for it took up to five days to travel from Dar es Salaam to the farthest outposts of administration.<sup>1</sup> Further pressure for roads and railways came from German plantation owners who had settled in the first years of the nineteenth century around Tanga and Moshi. The latter were linked by the Tanga-Moshi railway in 1911, while the Central Line from Dar es Salaam to Kigoma was completed in 1914. However, it was not to serve German ambitions: from 1914 to 1918, development was not only stopped but reversed. Roads, railways, telegraphs, and telegraph wires, all the space-linking innovations of the previous two decades, were sabotaged in the guerrilla warfare of von Lettow-Vorbeck, so that upon the assumption of British rule in 1919, administrative patterns and linkages had to be completely reestablished. For the most part, such renovations followed and consolidated those set up by the earlier German administration. The pattern of the early twenties forms the starting point for this

<sup>1</sup> Uhlig, *Erläuterungen zur Wirtschaftskarte von Deutsch-Ostafrika* (Berlin

analysis, and the first of five time-slices in which the structure of modernization is examined in its spatial context. Prior to such analysis, however, it is important to lay out clearly the data and methodological bases upon which the maps of modernization have been constructed.

#### DIRTY DATA, CLEAN GEOMETRIES, AND THE RESOLUTION OF SPATIAL LENSES

I shall not belabor the problem of data availability for such a study through time in Africa. When we examine the modernization process at a reasonable degree of spatial resolution, as well as through time, the constraints become unusually severe. For each time period, data were assigned to a hexagonal network of 289 cells covering the country (Figure 1), so that the geographer's basic requirement of spatial as-

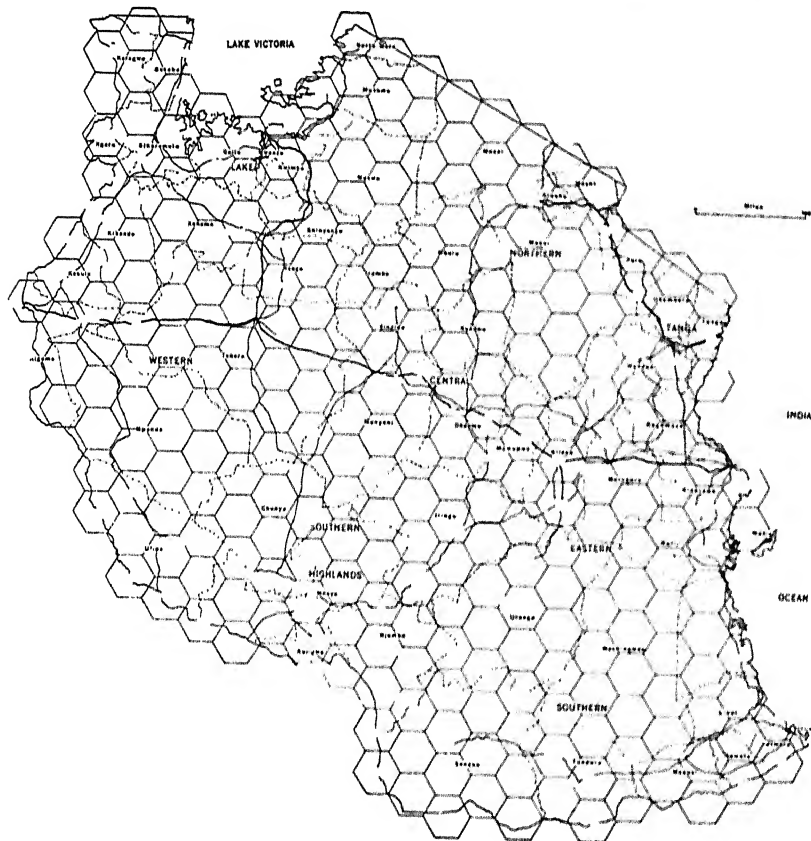


FIGURE 1. Hexagonal network of observation cells to which modernization data are assigned.

THI

signability  
content w  
gregation  
for areas v  
scale of c  
when it c  
square mi  
observatio  
of aggreg  
from one

The pro  
form, is a  
time perio  
they inde  
upon *both*  
units. I in  
severe assu  
tions for a

Each va  
sions may  
notions, s  
provide th  
complex n  
process we  
be roughl  
would be  
varying d  
tend to pe  
between t  
two variat  
coincident  
and seek  
whose dir  
tural patte  
be consid  
given by t  
project at  
then be us

<sup>2</sup> For an  
Analysis," C

signability imposes an immediate limitation. We can no longer be content with regional, provincial, or other high levels of spatial aggregation in which single, and normally useless, values are reported for areas varying up to tens of thousands of square miles. Rather, our scale of observation is such that information becomes usable only when it can be plausibly assigned to equal cells of approximately 125 square miles. Because of the limitation imposed by the basic units of observation, and because methods of statistical reporting and degrees of aggregation change over time, comparability of all the measures from one time period to another is impossible to achieve.

The problem of structuring such data expressed in standardized form, is approached from the purely geometrical viewpoint.<sup>2</sup> For each time period, a number of variables are available, and we assume that they index the spatial intensity of modernization. They are measured upon *both* interval and nominal scales over the range of observational units. I invoke no inferential notions whatsoever, since none of the severe assumptions holds, and we are seeking useful structural descriptions for a total population.

Each variable is considered as a unit vector in a space whose dimensions may be as numerous as the variables themselves. Such geometrical notions, so closely allied to the geographical, are important, for they provide those mental pictures that are essential whenever we discuss complex multivariate structures. If each measure of the modernization process were quite independent of the others, then each vector would be roughly at right angles to the rest, and the cosines between them would be close to zero. On the other hand, if the measures index in varying degrees a smaller number of underlying concepts, they will tend to point toward a few directions in the space, so that the cosines between them approach unity, a value obtained in practice only when two variables measure exactly the same thing so that their vectors are coincident. We may term the cosines correlation coefficients if we wish, and seek within the vector clusters particular axes called eigenvectors, whose direction and thrust in the space are determined by the structural pattern of the variables. Such principal components or axes may be considered as general synthesizing scales. Their interpretation is given by the degree to which the unit vectors representing the variables project at right angles upon them. Such projections, or loadings, may then be used as a set of weights to score each observation or hexagonal

<sup>2</sup> For an excellent elementary view, see Rudolf Rummel, "Understanding Factor Analysis," *Conflict Resolution*, xi (December 1967), 444-80.

cell. It is upon the basis of such scores that I have constructed the maps of the modernization surfaces (Figures 2-6).<sup>3</sup>

In addition to providing a method of combining and structuring multivariate data, the principal components approach, embodying least-squares criteria, serves as a filter for data that can only be characterized as noisy. There is, of course, error in the data—perhaps as high as ten percent in the case of one or two variables—but it seems quite possible that the error is unbiased and randomly distributed, serving only to *lower* relationships rather than the reverse.<sup>4</sup>

Unfortunately, error-reducing methods can do little about the lack of data. Among the information gaps, that indexing the location and intensity of education through time is by far the most serious. Except at totally useless levels of aggregation, I have been unable to find reasonable data on the locations or enrollments of the mission schools, which totaled 1800 by 1914—although dropping to 300 in 1934. Data for government schools are available, but these form such a minute proportion of the total that their inclusion would only serve to mislead. While we must deplore the omission of an educational index, there is evidence that schools, and school enrollments, have frequently gone hand in hand with other measures of the modernization process. Thus the gross patterns, and our insights derived from them, will not be greatly affected. Finally, variables have not been measured on a per capita basis. Early censuses were notorious for their inaccuracies, and I feel that it would be foolish to inject additional noise into the data sets when many of the subsets are relatively clean.

#### THE EARLY TWENTIES: CONSOLIDATION AND REPAIR

For the base period of the early nineteen-twenties, seven variables are available to index the spatial variation in the intensity of the modernization process (Table 1).<sup>5</sup> The number of road-miles "fit for light-wheeled vehicles" per hexagonal cell (1) contained in this time period, as in all others, the greatest error. Road qualities are seldom precisely defined or recorded, and where dirt roads are concerned it is obvious

<sup>3</sup> I have eschewed rotational schemes for the component structures to criteria formalizing the intuitions of psychologists for other purposes. In this context, I can offer no plausible argument for the usual normalized varimax criterion, which appears simply as an inefficient taxonomic device for which a classification of variables is achieved under a severe and quite unrealistic orthogonal constraint.

<sup>4</sup> In the context of historical studies see David McClelland, *The Achieving Society* (New York 1961), 61.

<sup>5</sup> Since I have used huge data matrices up to 290x26, I report only tables of component loadings.

| TABLE I - COMPONENT LOADINGS  |                |      |                                  |      |               |      |                                |      |                                |      |
|-------------------------------|----------------|------|----------------------------------|------|---------------|------|--------------------------------|------|--------------------------------|------|
| Variables                     | Early Twenties |      | Late Twenties-<br>Early Thirties |      | Late Thirties |      | Late Forties-<br>Early Fifties |      | Late Fifties-<br>Early Sixties |      |
|                               | I              | II   | I                                | II   | I             | II   | I                              | II   | I                              | II   |
| 1. Total Road Miles           | 47             | -69  | 54                               | 24   | 56            | 45   | 83                             | 26   | 78                             | 40   |
| 2. Administrative Officers    | 91             | 06   | 88                               | -09  | 95            | -16  | 73                             | -54  | 68                             | -69  |
| 3. Police                     | 85             | 02   | 88                               | -32  | 93            | -01  | 91                             | -35  | 88                             | -47  |
| 4. Railway                    | 29             | 75   | 41                               | -11  | 44            | 32   | 46                             | 14   | 42                             | 22   |
| 5. Hospital Beds              | 92             | 07   |                                  |      |               |      | 96                             | -18  | 95                             | -02  |
| 6. Piped Water                |                |      | 75                               | -21  |               |      | 71                             | 50   | 61                             | 44   |
| 7. Telegraph                  |                |      | 80                               | 10   |               |      | 75                             | 40   | 68                             | 43   |
| 8. Telephone                  |                |      | 69                               | -29  | 76            | 30   | 76                             | 44   |                                |      |
| 9. Post Offices               |                |      | 74                               | 35   | 65            | 43   |                                |      | 85                             | 18   |
| 10. Criminal Convictions      |                |      | 88                               | -10  | 88            | 14   |                                |      | 94                             | -22  |
| 11. Electricity               |                |      | 62                               | -33  |               |      | 77                             | 18   | 72                             | 26   |
| 12. Other Government Officers |                |      |                                  |      | 85            | -46  | 80                             | -53  | 90                             | -38  |
| 13. High Court Circuits       |                |      |                                  |      | 93            | -01  | 86                             | -35  | 85                             | -25  |
| 14. Banks                     |                |      | 83                               | -43  |               |      | 88                             | -25  |                                |      |
| 15. African P.O. Savings      |                |      | 82                               | 07   |               |      |                                |      | 84                             | 21   |
| 16. Unskilled Wages           |                |      | 88                               | 29   |               |      | 69                             | 26   |                                |      |
| 17. Tribal Dispensaries       |                |      | 20                               | 48   | 18            | 64   |                                |      |                                |      |
| 18. Total Eco. Functions      |                |      |                                  |      | 83            | -48  |                                |      | 85                             | -51  |
| 19. Govt. Medical Personnel   |                |      |                                  |      |               |      | 80                             | -56  | 81                             | -56  |
| 20. Airfield Quality          |                |      |                                  |      |               |      | 80                             | 39   | 71                             | 43   |
| 21. Police Radio              |                |      |                                  |      |               |      | 65                             | 11   | 87                             | -05  |
| 22. Agricultural Officers     | 72             | -30  |                                  |      |               |      |                                |      |                                |      |
| 23. Sanitary Staff            | 72             | 20   |                                  |      |               |      |                                |      |                                |      |
| 24. Govt. Hospitals           |                |      | 74                               | 46   |               |      |                                |      |                                |      |
| 25. Designated Township       |                |      | 74                               | 40   |               |      |                                |      |                                |      |
| 26. Haulage License Appins.   |                |      |                                  |      |               |      |                                |      | 85                             | 08   |
| 27. Commercial Estates        |                |      | 71                               | -16  |               |      |                                |      | 68                             | 45   |
| 28. Development Loans         |                |      |                                  |      |               |      |                                |      |                                |      |
| 29. Criminal Cases            |                |      |                                  |      |               |      | 85                             | -37  |                                |      |
| 30. Doctors                   |                |      |                                  |      | 95            | -19  |                                |      |                                |      |
| 31. Quality Stores            |                |      |                                  |      |               |      | 86                             | 38   |                                |      |
| 32. Churches                  |                |      |                                  |      |               |      | 78                             | 38   |                                |      |
| 33. Cinemas                   |                |      |                                  |      |               |      | 82                             | 03   |                                |      |
| 34. Prison Population         |                |      |                                  |      |               |      | 78                             | -03  |                                |      |
| 35. Mission Hospitals         |                |      |                                  |      |               |      |                                |      | 28                             | 26   |
| 36. Teacher Training          |                |      |                                  |      |               |      |                                |      | 62                             | 31   |
| 37. Secondary Schools         |                |      |                                  |      |               |      |                                |      | 65                             | 18   |
| 38. Middle Schools            |                |      |                                  |      |               |      |                                |      | 63                             | 48   |
| 39. Bituminized Roads         |                |      |                                  |      |               |      |                                |      | 73                             | 00   |
| 40. Telephone Subscribers     |                |      |                                  |      |               |      |                                |      | 80                             | -58  |
| 41. Asian Population          |                |      |                                  |      |               |      |                                |      | 85                             | 21   |
| Eigenvalues                   | 4.02           | 1.18 | 9.17                             | 1.47 | 7.28          | 1.50 | 13.15                          | 2.61 | 15.14                          | 3.46 |
| Percent Total Variation       | 57.4           | 16.9 | 53.9                             | 8.6  | 60.7          | 12.5 | 62.6                           | 12.4 | 58.2                           | 13.3 |
| Number of Variables Available | 7              |      | 17                               |      | 12            |      | 21                             |      | 26                             |      |

that conditions vary markedly from season to season. I have used considerable subjective judgment in the definition of such roads, and we must be somewhat cautious where this variable enters the analysis. The railway (4) is simply scaled as an attribute, since the network, or tree, remains rudimentary in form and the sheer presence or absence in a cell is all we require to measure. The number of administrative officers (2) indicates the locational intensity of the overall administrative impress. Ideally we could refine such figures by recording the number and routes of safaris by District Officers, but such data are available only in the old district diaries—many of which appear to have been lost or stolen. European agricultural officers and African agricultural instructors (22) indexes the availability of new agricultural practices and the possibility of agricultural innovation, while the number of police (3) indicates the intensity of the ultimate power to enforce political decisions and control. Finally, the number of hospital beds (5) and the number of African sanitary staff (23) measure the availability of modern medicine and hygiene.

Turning to our geometrical model and conceptual image, we see that each of the unit vectors representing these variables is far from orthogonal from the others in the seven-dimensional space. Indeed, there is a strong directional thrust to five of the variables, measured by the loadings on the first component (Table 1). Such a major component, accounting for nearly sixty percent of the variation, may be interpreted as a scale measuring the major aspect of the modernization process. Note, however, the way in which both of the transport and communication variables document a second dimension at this time. Because their space-bridging, as opposed to space-organizing, functions were so critical at this time, both road and rail appear only weakly related to the other measures, forming a distinct, though less important dimension to the modernization process.

If we calculate scores on the major scale of modernization for each cell and plot them cartographically, we can construct a modernization surface for the early twenties over Tanzania on which heights are always relative to the greatest value of 1,000 at Dar es Salaam (Figure 2). The phrases "Islands of Development" and "Islands of Modernization" are frequently used in the African context, but such verbal images are sharply emphasized by such a map. Only a little over a third of the country was touched, let alone impressed, by the process, and even this is an over-generous estimate, for only the presence of roads serving as bridges between small nodes allows a large number



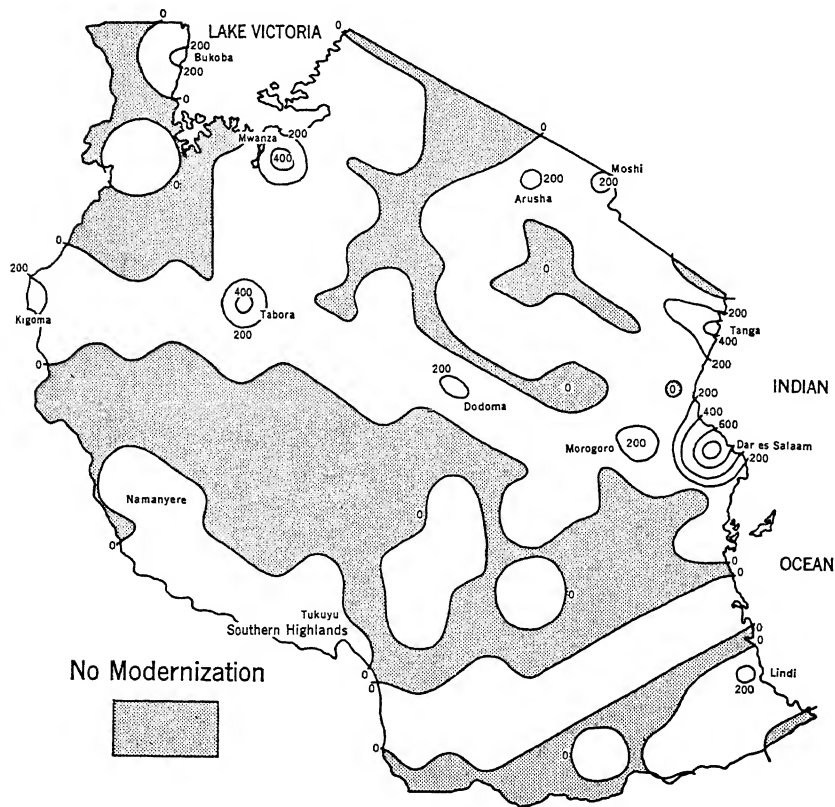


FIGURE 2. Tanzania: The modernization surface in the early twenties.

of the cells to be scored. Indeed, it is difficult to speak of a single surface at all; we have rather fragmented and disconnected pieces that appear to float in the still mill-pond of traditionalism (shaded pattern). The cell containing Dar es Salaam (1,000) dominates the surface, rising roughly twice as high as Mwanza (563), Tabora (515), and Tanga (408). The latter has yet to recover from the effects of the war, and in its rail hinterland Moshi (277) and Arusha (269) are but small intrusions forcing their way up to recover their pre-war positions. The Central Railway forms a distinct axis on the map, with small nodes at Morogoro, Dodoma, and Kigoma. South of this line no value on the surface reaches 200, with the exception of Lindi in the southeast. The Southern Highlands, remote and virtually inaccessible by modern transport at this time, have yet to show a distinct



peak, although small detachments of police and rudimentary hospital facilities could be found at the administrative outposts of Tukuyu and Namanyere. To the north, on the western shore of Lake Victoria, Bukoba emerges as a small island with one-fifth of the intensity of Dar es Salaam.

THE LATE TWENTIES AND EARLY THIRTIES:  
FIRST THRUST FROM THE CONSOLIDATING YEARS

Although much of the criticism of Governor Byatt in the early twenties was grossly unfair, in view of the enormous difficulties he faced at the time of mandate, there is little doubt that the appointment of Cameron in 1925 brought a fresh point of view, new energy, and an air of optimism to the task of development. With its emphasis on communications and indirect rule, with commercial agriculture expanding rapidly to form a financial base, his era was marked by a surge of activity unequalled until the late forties. Such a thrust is reflected by the changes in the modernization surface (Figure 3), constructed from weighted combinations of seventeen variables (Table 1).

The overall orientation of the vectors is still marked, with the first component accounting for fifty-four percent of the variation. Once again, only the transport variables Road Miles (1) and Railway (4), together with Tribal Dispensaries (17), load low, but both have increased over the first period. The increases are slight, and our interpretation must be cautious in view of the error component in the Road Miles, but there is some slight evidence that the other measures of modernization are becoming more strongly related to these fundamental space-linking and space-organizing variables. We must also ascribe to the tribal dispensaries an importance never emphasized before. Catalyzed by the great energy and enthusiasm of the medical officers in the early mandate years, and supported by native treasury funds, hundreds of small dispensaries and dressing stations were in operation by 1929, and by 1934 were treating half a million patients per year.<sup>6</sup> Thus, for many rural Tanzanians, roads and medical dispensaries were the first aspects of the modernization process to appear with any permanence in their areas. In a very real sense, they appear to be lead variables, a small group appearing before all other aspects,

<sup>6</sup> J. P. Moffett, ed., *Handbook of Tanganyika* (Dar es Salaam 1958), 111.

and so in  
total mea

Unlike  
cation me  
but comb  
nent to t  
severely c  
another i  
between t  
break in  
ing capac  
as a weak  
tribal disp  
offices (9  
with ban  
of a still

The ch  
sixty-five  
formerly  
pockets. I  
rather th  
Road-buil  
also thick  
The latter  
Dar es Sa  
Tanga, w  
Mountain

<sup>7</sup> The que  
frequently at  
Gould, *The*  
106-14; also  
in Underde  
4, 528.

<sup>8</sup> Raymon  
of *Multivari*

<sup>9</sup> While I  
eigenvalues  
trast to  $\lambda_1/\lambda$   
through the  
ellipsoidal c  
circularity, s  
ness of locat  
and error. F  
actions of th

and so initially poorly related to the remainder of the items in the total measurement set.<sup>7</sup>

Unlike the case in the earlier period, the transport and communication measures no longer define a distinct and separate dimension, but combine with other variables to index a possible second component to the modernization process. Such an interpretation must be severely qualified, for the evidence that the second component forms another independent scale is slight. In the first place, the disparity between the first and second eigenvalues is large,<sup>8</sup> and there is a clear break in the sequence of these measures describing "variance absorbing capacity." Nevertheless, I shall interpret the second component as a weakly emerging scale contrasting measures such as hospitals (24), tribal dispensaries (17), official township designation (25), and post offices (9) that are found in rural areas and smaller central places with banks (14) and electricity (11) that are exclusively attributes of a still rudimentary urban system.

The change in the modernization surface (Figure 3) is striking; sixty-five percent of the cells enter the analysis, and the large areas formerly untouched by the surface have been squeezed into smaller pockets. Despite such gaps, we can truly speak of a surface now, rather than of the series of fragmented pieces of the earlier period. Road-building activity has not only linked the bits together, but has also thickened the networks focusing upon the major urban nodes. The latter, however, still define distinct peaks, the greatest being at Dar es Salaam, against which all the others are expressed relatively. Tanga, with its rich hinterland in the coastal plain and the Usambara Mountains, has recovered to take second place with a score of 888,

<sup>7</sup> The question of lag and lead is a difficult one to disentangle, since there are frequently strong interdependencies between transport and other phenomena. See Peter Gould, *The Development of the Transportation Pattern in Ghana* (Evanston 1960), 106-14; also Edward Taaffe, Richard Morrill, and Peter Gould, "Transport Expansion in Underdeveloped Countries: A Comparative Analysis," *Geographical Review*, LIII, 1963, 528.

<sup>8</sup> Raymond Cattell, "The Meaning and Strategic Use of Factor Analysis," *Handbook of Multivariate Experimental Psychology* (Chicago 1966), 206.

<sup>9</sup> While I have not seen such an argument in the literature, the ratios of successive eigenvalues quickly approach unity after the first component, with  $\lambda_2/\lambda_3 = 1.3$  in contrast to  $\lambda_1/\lambda_2 = 6.2$ . If we think of the eigenvalues as "stretchability coefficients," a cut through the hyper-figure along the major axis indicates a strong directional thrust and ellipsoidal configuration. But successive cuts orthogonal to the first indicate only rough circularity, so that the locations of the other axes are virtually arbitrary. Such arbitrariness of location may indicate that we are absorbing the irrelevant variance of uniqueness and error. Peter Gould, "On the Geographical Interpretation of Eigenvalues," *Transactions of the Institute of British Geographers*, No. 42 (1967) 58.

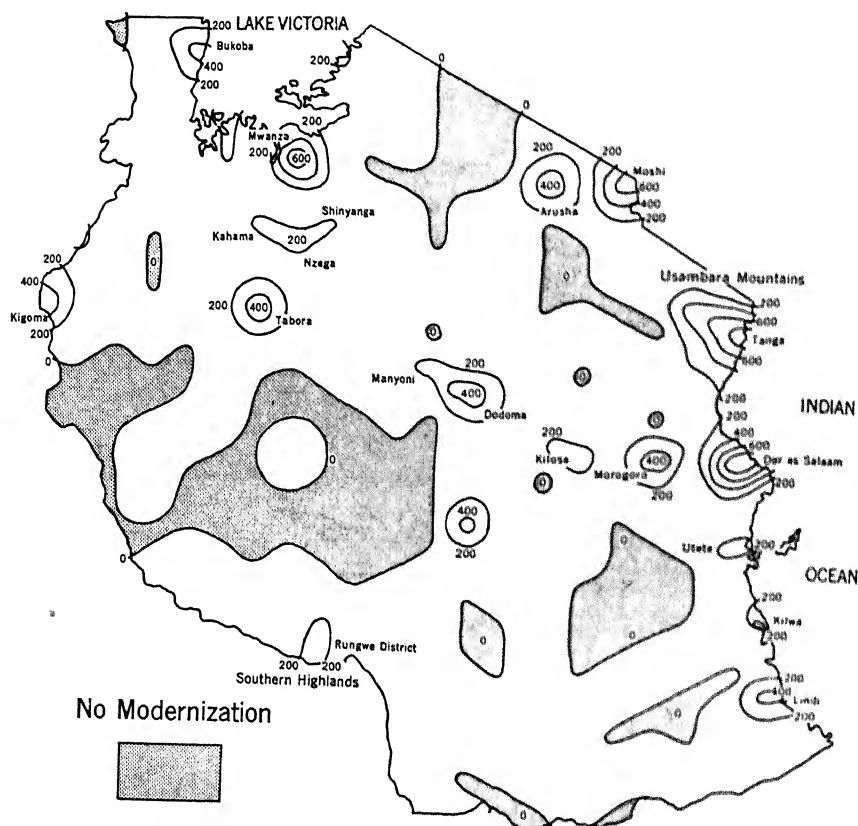


FIGURE 3. Tanzania: The modernization surface in the late twenties and early thirties.

and the area enclosed by the 200 isoline exceeds that of Dar es Salaam itself. The latter peak has extremely steep gradients, implying that once the city is left the decline in modernization is rapid. Little appears to have "trickled down" from this major administrative center to the surrounding hinterland. Indeed, such steep gradients characterize all of the other nodes. Moshi (656) is third, followed by Mwanza (647) and Arusha (585), but such intensities also appear extremely local in character. Along the Central Railway axis, Tabora (577) has gained only slightly from the previous period, but Dodoma (479) has risen rapidly, reaching west along the railway lines to an outlier at Manyoni (285). Morogoro (537) and Kigoma (501) are also sharp intrusions, and for the first time a new peak appears at Kilosa (284) between previously established nodes. The emergence of nodes of moderniza-

tion approximately halfway between established ones also characterizes the Shinyanga-Nzega-Kahama belt (201, 267, and 215) between Mwanza and Tabora along the newly-completed rail line to Lake Victoria; the Iringa node on the Great North Road between Dodoma and Rungwe District in the Southern Highlands; and the Kilwa (307) and Utete (262) nodes between Dar es Salaam and Lindi (451). The latter has risen sharply, and while the blank areas of its hinterland have been greatly reduced to smaller pockets a vast area of the south remains at an extremely low level.

As a whole, and relative to Dar es Salaam at this time, the modernization surface conveys an impression of considerable change from the last. Though discernible earlier, most of the urban growth poles appear far more prominent as many aspects of the modernization process converge to these towns that provide the foci for small nodal or functional regions round about them. Note once again, however, that gradients are extremely steep in all directions. To use the language of geomorphology, we still have inselbergs of modernization on a low-lying plain that is barely above tradition-level.

#### THE LATE THIRTIES: RECOVERY FROM DEPRESSION

Triggered by circumstances thousands of miles away and beyond the control of Tanzania, the Depression of the early thirties forced cuts in virtually every aspect of the development and modernization process. By the middle of the decade it became more a matter of keeping what had already been established, rather than of extending the surge of the previous ten years. However, by the end of the decade, and despite political uncertainties, most financial indices—exports, budgetary allocations, etc.—were back to previous pre-Depression levels.

Only twelve measures are available for this period (Table 1), but an even stronger overall thrust is apparent with sixty-one percent of the variation being accounted for by the first component. The road and rail variables maintain their loadings, although these should still be interpreted cautiously, for their contribution to the major dimension becomes somewhat more complex. Road Miles (1) and Railway (4) also contribute to the second scale, together with Tribal Dispensaries (17) and Post Offices (9). As these are in contrast to Total Economic Functions (18) and Other Government Officers (12), we have further evidence for the appearance of a second relevant dimension in which essentially rural attributes are contrasted to those found mainly in the

urban areas. Thus Road Miles (1) now contribute both to a major dimension of modernization and to a weak but emerging minor scale contrasting urban-rural aspects. They do this by virtue of their space-organizing role, focusing the immediate hinterland areas upon the urban nodes, as well as their space-bridging or linking role that locates them in essentially rural areas still untouched by many other aspects of the modernization process.

When interpreting the modernization surface constructed from the scores on the first component (Figure 4), we must remember that it has been normalized with respect to the highest peak—Dar es Salaam. Thus comparisons to previous surfaces must be made with care, for a sudden, jolting surge in the capital can lower the other areas when they are scaled in relation to it. The contrasts are vivid, nevertheless, and indicate that in these relative terms the primacy of the capital increased substantially. Tanga, still second with 469, cannot reach half-

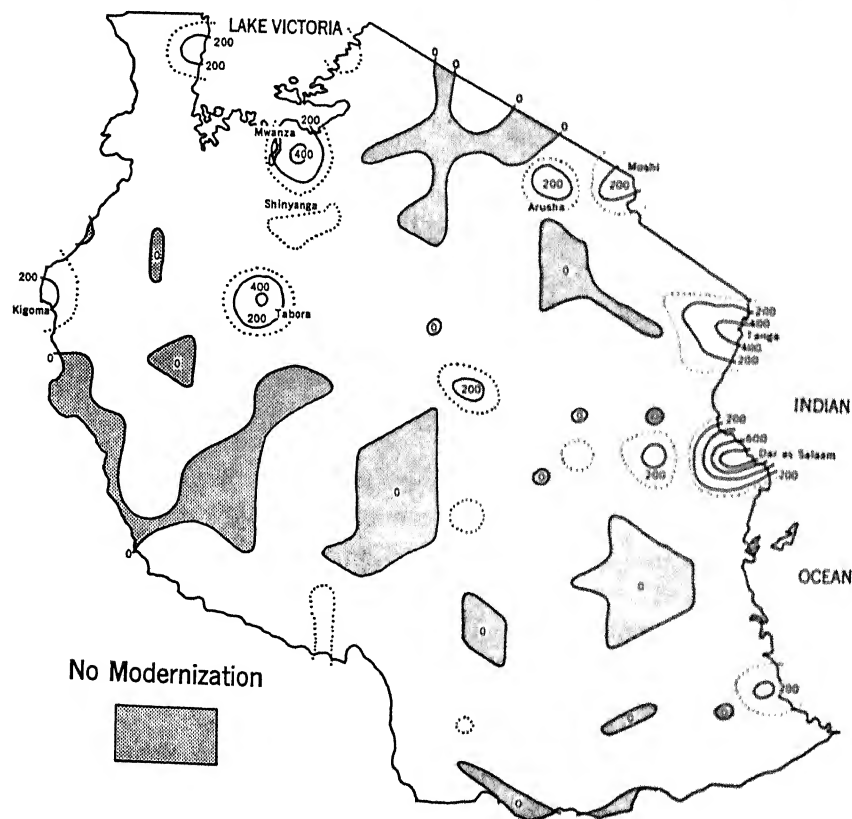


FIGURE 4. Tanzania: The modernization surface in the late thirties.



way to Dar es Salaam's peak, although it had a score of 888 ten years before. Mwanza (455) and Tabora (419) have also lost ground, though they are now themselves relatively larger than the Shinyanga node (118), which no longer appears as a distinct peak.<sup>10</sup> Arusha (308) and Moshi (307), are virtually equal, but the trend for the future is already in Arusha's favor. The greatest decline in all has been suffered by Kigoma (223), at the far western end of the Central Railway. As it is literally "out of things," and with its surrounding area never fulfilling the hopes of earlier years, the few measures that are directly comparable between time periods indicate that the town has sunk back in *absolute* as well as in relative terms. Elsewhere the process of "space-pinching" has proceeded, though at a far slower pace than in the twenties. Nearly seventy percent of the 289 cells enter the analysis, and the remaining blank pockets are practically devoid of population.

#### LATE FORTIES-EARLY FIFTIES: THE POSTWAR SURGE

In the same way that the depression of the thirties thwarted the growth of the twenties, so the Second World War cut short the period of slow but steady recovery. After the war, however, the assumptions and attitudes of many trusteeship administrators differed considerably from those of their prewar colleagues. Many of the more sensitive felt the Winds of Change in the first postwar years, and planning horizons to the day of independence, while still seen dimly, shortened considerably from the vague estimates of a decade before. Furthermore, the attitudes of a small but growing group of Tanzanians began to alter, so that by the late forties and early fifties the pace of development began to be accelerated by views and values that had hardly constituted a force before.

Twenty-one measures of the modernization process are available (Table 1), with the first two major components accounting for sixty-three percent and twelve percent of the variation—figures virtually identical to those of the late-thirties. One striking difference should be noted however: whereas the road variable (1) contributed to both components in the immediate pre-war years, it now loads strongly on the first, contributing at this time only to the major dimension of modernization. With their space-linking task completed by the war, the roads now take on much more strongly a space-organizing and space-drain-

<sup>10</sup> Note that the 100 isoline has been included as a dotted line to identify the relatively weaker nodes of this period.

ing function. Thus feeder densities thicken in the areas of expanded economic development—areas that focus, in turn, upon the modernization peaks and growth poles that assume to an ever greater extent the role of regional foci.

Nearly all the variables load highly and contribute to the major scale, indicating that modernization continues to be a strongly integrative process in which virtually all measures go hand in hand. No matter which aspect we choose—administrative, medical, judicial, economic, or communications—all are reflected in the overall surge. Only the railway variable (4) loads weakly, defining a specific dimension almost by itself later in the sequential component analysis. Though a vital form of transportation upon which most of the export flow depends for its journey to the ports, the railway appears in areas of intense growth as well as in others almost bereft of modern attributes. Thus its overall relationship to other measures remains weak. Unlike the roads, with their cheap and flexible space-searching ability, the railway demands huge capital lumps, so that it is spatially inflexible and quite unresponsive to many forms of local transport demand that reflect modern developments in a regional, as opposed to national, context.

Despite a lack of the “obvious” rural measures playing the same sort of key or indexing role as tribal dispensaries did in the late thirties, the second dimension may nevertheless be cautiously interpreted as a later manifestation of the rural-urban dimension that started to emerge in the earlier period.<sup>11</sup> Most government functions (administrative, medical, and judicial), are concentrated in the towns, while the variables indexing and serving the private sector are not confined to the urban centers.

The modernization surface (Figure 5) reflects the remarkable changes that had come about during the first postwar period. While roughly the same number of cells enter the analysis, small changes appear in the shape of the empty areas as little-used road-links are abandoned and others added. The abandonment of the rough dirt road from Mahenge to the south is a typical indication of the weakness of north-south internal links and the general lack of interaction between the interior portion of the Southern Province and other parts of the country.

<sup>11</sup> With a  $\lambda_2/\lambda_3$  ratio of 2.5, stronger evidence is now available for a bipolar component contrasting government functions (2, 3, 12, 13, 29, and 30), with those more generally in the private sector (6, 16, 31, and 32) together with the communications variables serving it (7 and 8).

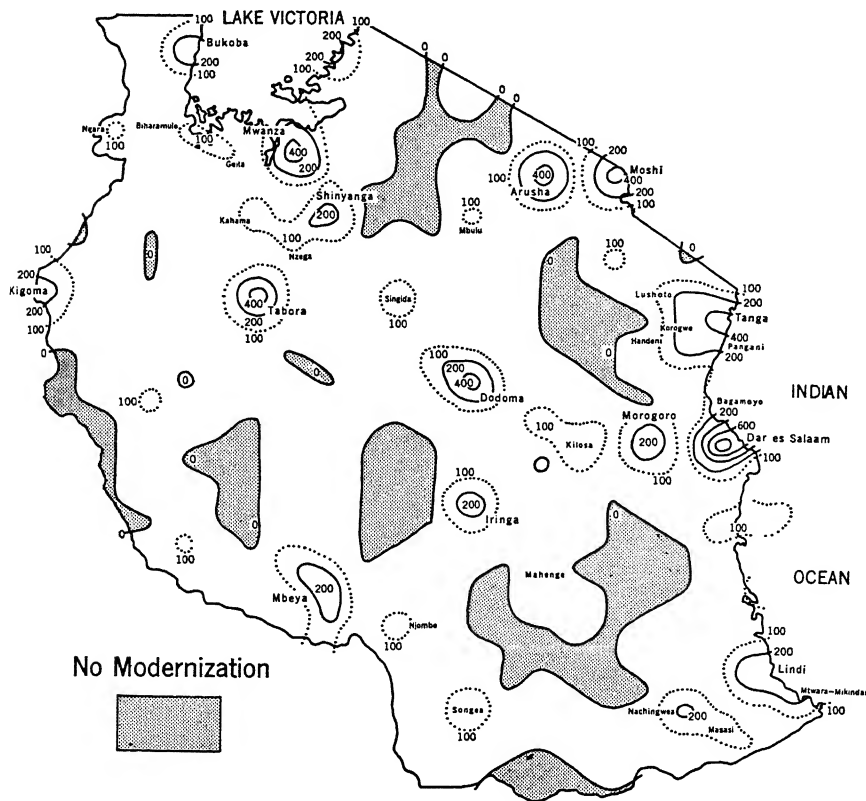


FIGURE 5. Tanzania: The modernization surface in the late forties and early fifties.

While strict and absolute comparisons are difficult, the general pattern resembles that of the late twenties-early thirties (Figure 3), providing some indication of the general stability and logical development of spatial forms inherent in the modernization process. Nevertheless, to use a photographic analogy, the latent image of modernization is far more developed by the late forties-early fifties. Particularly striking are the changes in the south as the ill-fated Groundnut Scheme sends a jolt of development through the area. The Lindi and Mtwara-Mikindani nodes have coalesced, while an entirely new node at Nachingwea appears to couple with the older and far weaker node at Masasi. The Songea area has also expanded slightly, and between it and the major peak of Mbeya in the Southern Highlands the Njombe node just begins to emerge. Indeed, the geographical characteristic of



"betweenness" appears as a dominant theme at this time. In the west, the Sumbwanga and Mpanda nodes break through the one-hundred isoline and the plain of traditionalism for the first time (dotted line), while north of the railway the Ngara, Geita-Biharamulo, Singida, Mbulu, and Same nodes all seem to fill in gaps in the spatial lattice. Along the railway, postwar developments have expanded the Kilosa node to join with Mpwapwa, though interestingly this less concentrated pattern continues to be dominated by the Morogoro, Iringa, and Dodoma peaks that surround it. On the railway line between Tabora and Mwanza, the Shinyanga-Nzega-Kahama belt emerges strongly again, although the eastern portion, centered on Shinyanga, dominates this somewhat smeared concentration of modern elements. In the Northern Province the Gemini of modernization—Arusha and Moshi—continue their tradition of rivalry. Finally, the Tanga hinterland, linking the Lushoto, Korogwe-Handeni, and Pangani nodes, forms the only area where there is strong evidence for a "trickle-down" effect from a major modernizing center to its surrounding area. Dar es Salaam dominates the country as usual, but the steep gradients from the peak indicate the extremely local effect of the capital city upon its immediate hinterland.

#### LATE FIFTIES-EARLY SIXTIES: THE INDEPENDENCE YEARS

In the final time-slice of the analysis around the years of Independence, twenty-six measures are available (Table 1). In addition to a number of variables used previously, there are also some indexing educational facilities (33, 36, and 37) and the economic service functions performed by the towns (18). I have deliberately included Asian population (41), for we have strong evidence that this population-group was a potent surrogate measure of local trade in East Africa at this time.<sup>12</sup>

Once again, the familiar spatial and structural patterns emerge, with all measures loading highly on the first component with the exception of the railway (4) and mission hospitals (35). What is of particular interest, however, is the continuing strength of the second scale, which accounts for approximately thirteen percent of the variance.<sup>13</sup> Thus the basic structure of the modernization process appears to be a major scale of considerable stability through time, to which virtually all indices contribute highly, together with a second, urban-rural scale that

<sup>12</sup> Edward W. Soja, *The Geography of Modernization in Kenya* (Syracuse 1968), 78, 92-93.

<sup>13</sup> The ratio of the second and third eigenvalues is now 2.60.

becomes more significant as time proceeds. Upon the latter, all government (2, 3, 12, and 19), economic (18), and local communication (40) measures are contrasted against the variables indexing modern aspects of developing rural areas such as development loans (28), piped water (6), middle schools (38), and the less intense communications facilities, such as telegraph (7), roads (1), and airfields with good landing facilities (20), which increasingly supplement the poorer surface forms of transportation. Interestingly, the Asian population variable (41), having made a substantial contribution to the first component, shows by its sign that it is on the side of the modernizing rural indices, an indication of the presence of Asian trade not only in the large cities but also in the smaller central places and nodes of modernization.

Because the intense concentration of all aspects of modernization in Dar es Salaam alters relative contrasts, additional isolines of one hundred (large dot) and fifty (small dot) have been added to the surface to accentuate the smaller, emerging humps (Figure 6). Dar es Salaam still towers as a peak over its immediate area, but so steep are the gradients that it still cannot link with the Morogoro node to the west. The latter, however, has coalesced with the Kilosa node, which in turn linked with the Mpwapa area at an earlier time. Considering the map as a model in the very real sense of an abstraction, simplification, and compression of reality, we can predict the emergence soon of a national trunk or magistral as the Dodoma and Singida nodes expand and join the major penetration line from the capital. Shortly afterwards, and particularly in view of the Kilombero Valley developments based on a small rail spur south from Kilosa, the Iringa node will join, bringing the Sao Hill, Njombe, and Mbeye nodes in its train. In Lake Province the same process emerges as the Mwanza node links with the Shinyanga belt, and as the modernization process becomes more intense the spatial coalescence of this larger area to the Tabora and Singida nodes is only a matter of time. Similarly, in the northeast, the Arusha, Moshi, Same, and Tanga areas are reaching out toward one another along the railway to form a further continuous strip, and with the new rail link between the northern and central railways, it can be only a matter of a decade before this discontinuous ridge of modernization is joined to the main trunk.

The south, however, remains out of the general system. While the Nachingwea-Masasi axis has strengthened, it has yet to coalesce with the coastal foci of Lindi and Mtwara. Elsewhere, small and distinctly isolated nodes appear at Kigoma, Mpanda, Kibondo, Ngara, and Geita,

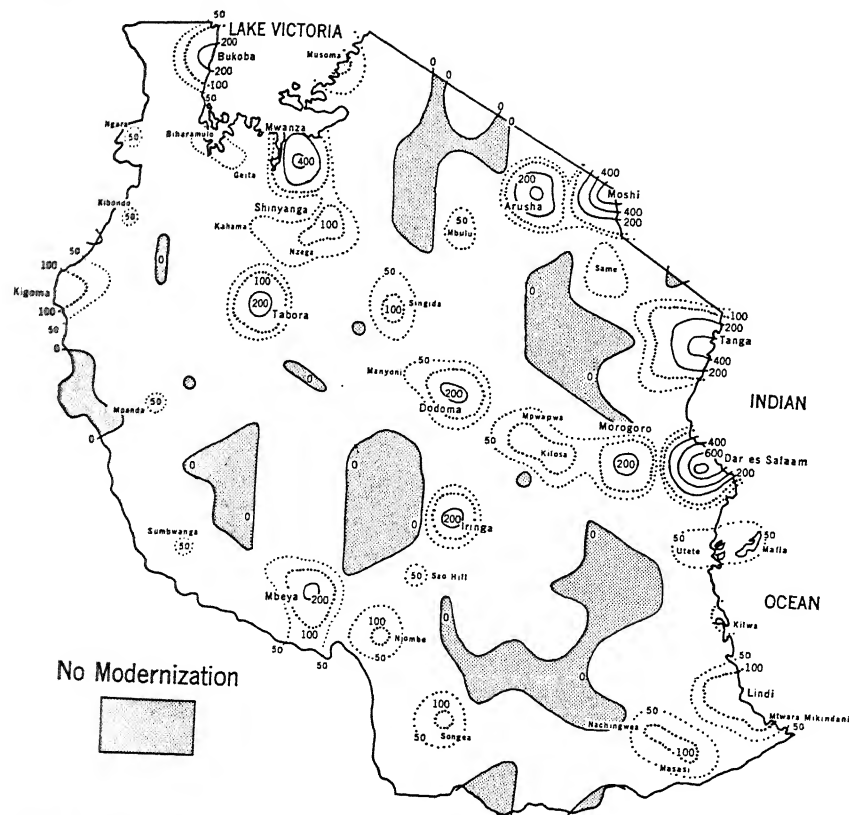


FIGURE 6. Tanzania: The modernization surface in the late fifties and early sixties.

although the latter may well serve in later years as the bridge that brings isolated Bukoba into the national system, linking it with Mwanza in a developed crescent around the western lake shore. In a similar way, Musoma, on the eastern side, may also link up, especially in view of the high quality "autostrada" road that has just been completed to the Mwanza railhead.

#### TENACITY OF PATTERN AND THE DYNAMICS OF CENTRAL PLACE SYSTEMS

Although we conclude on a somewhat predictive note, it would be foolish to imply that the analysis has illuminated the *process*, the actual dynamics of modernization in Tanzania. But what is clear, from the repetitive sequence of static patterns through time, is that whenever we consider modernization we examine to a very high degree a deeper

process of urbanization. Though it pushes the question back a step, such a statement is nothing new: writer after writer has noted the way in which "the town is the door through which Africa is entering the modern world."<sup>14</sup> But although it is intimately linked with modernization, what we normally refer to as urbanization is itself only an intense aspect of a more general and continuing process of central place development. To emphasize the intense degree to which a modernizing process is convoluted with the dynamics of urbanization, I have transformed the map of Tanzania from familiar geographic to "modernization" space (Figure 7). The areas of the hexagonal observation cells (Figure 1), are now proportional to their modernization scores. We lose shape, of course, but we have maintained all contiguity relation-

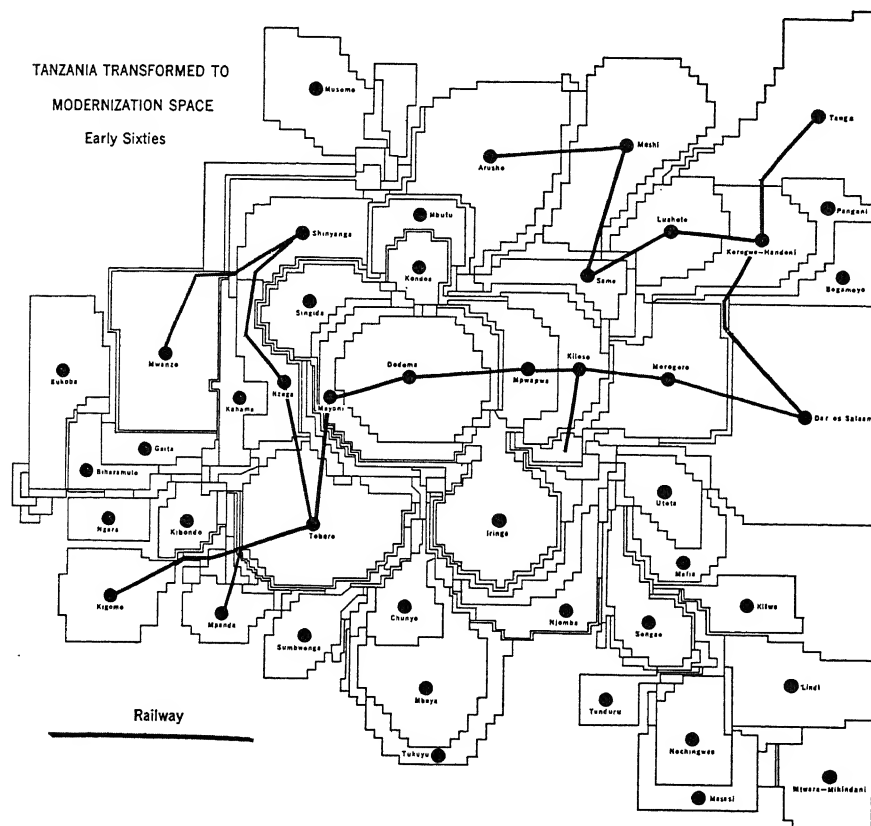


FIGURE 7. The map of Tanzania transformed from geographic to "modernization" space for the late fifties-early sixties.

<sup>14</sup>Daniel F. McCall, "The Dynamics of Urbanization in Africa," *Annals of the American Academy of Political and Social Science*, 298 (March 1955), 160.

ships. Despite the extreme distortions, the transformation gives a better view of the spatial aspects of modernization than does the more familiar map. Particularly striking is the dominance of the urban nodes and the collapse of the interstitial areas. When we speak of modernization we are clearly discussing an emerging and strengthening urban system. In addition, the weakness of the south is highlighted, emphasizing once again its developmental lag.

Thus, what we have seen emerging in the space called Tanzania between 1920 and 1963, and underlying the very modernization process we have tried to measure and structure, are the dynamics of that most spatial of systems—the central place structure; a *system* of *nodes* of various *sizes*, at differing *distances*, *linked* with varying intensities, influencing areas *contiguous* and *between*, structuring, focusing, and serving their hinterlands, and acting as emergent poles of attraction for the surrounding population. For a crucial element in the development of central places is a rural-to-urban migration. In the African context this always means the breaking of old allegiances to extended family, traditional authority, unquestioned belief, own people, and familiar place, and the forging of new allegiances in the very process of questioning the old. And so, underlying any theory of modernization in space and time, there must be a deeper theory of the dynamics of central-place systems.

Such a theory we do not have. Except for a few Monte Carlo simulations of developments in nineteenth-century Europe,<sup>15</sup> our theory is static and geometric, virtually ignoring the concomitant development of transportation and communication linkages, the very channels that allow the time-frozen *patterns* in space to become *systems* of human interaction and development. But there is increasing doubt that disaggregation of the spatial system into static patterns of nodes, linkages, and movement will lead to the writing of illuminating theory, for every piece of empirical evidence we have confirms the tightly interlocked nature of all these aspects.<sup>16</sup>

<sup>15</sup> Richard L. Morrill, *Migration and the Spread of Urban Settlement*, Lund Studies in Geography, Series B, No. 26 (Lund 1965).

<sup>16</sup> Telephone flows among modernized nodes in Tanzania, for example, are predictable to a high degree by relating them in simple gravity-model form to the size of the modernization scores of the generating and receiving centers and the distance between them ( $R^2_{1,234} = 0.58$ ). This can be expressed in standardized regression coefficients after logarithmic transformation:

$$X_{c1,234} = 0.4524X_2 + 0.4332X_3 - 0.6536X_4$$

|                    |                                 |                                |                            |
|--------------------|---------------------------------|--------------------------------|----------------------------|
| Telephone<br>Calls | Score of Generat-<br>ing Center | Score of Re-<br>ceiving Center | Distance Between<br>Center |
|--------------------|---------------------------------|--------------------------------|----------------------------|

It is hardly surprising, then, that a concern for systems analysis is permeating every social and behavioral science today, for only such an approach and framework for thought can hope to achieve an understanding of the dynamics of human society. Simply consider Tanzania in 1920, its people unevenly spread in clusters as the land, the rain, and the tsetse fly (themselves intimately related) allow more in some areas than others. Administrators are sent by the early colonial and mandate powers to the population clusters, not the empty areas, and the administrators themselves must move and communicate. Roads and railways link the administrative nodes and provide, in turn, channels through which modernizing innovations seep. From 1920 to 1963, the road vector loads 0.47, 0.54, 0.56, 0.83, and 0.78—ever higher upon the main dimension of modernization. But this is a result of a mathematical requirement of the model: what is happening, of course, is that the major cluster of modernizing variables is moving ever closer to the single road vector, as in geographic, rather than geometric, space, modernization in all its innovative aspects is distributed from the major sources through the tarred arteries and laterite capillaries of the land and society. Some innovations have economic implications—coffee revolutionizes Kilimanjaro, sisal spreads around Tanga and Morogoro, cotton seeps through Sukamaland—and commodity flows swirl through the road network, feeding back information to the administrators who upgrade, realign, and tar the dirt tracks of the previous year. Other modernizing activities swell the small towns, and these in turn attract the surrounding people. Such human movements and interactions themselves become flows of messages feeding back new ideas and innovations to the villages. Roads and central places grow to serve the population, but the very pattern of the people is itself pulled and warped through time into greater conformity with the central-place structure and its network of linkages. Where demands of all sorts grow—administrative, medical, educational, economic—response is called forth. And so it is that the early patterns of the twenties (Figure 2), are still clearly discernible in the pattern at the time of independence (Figure 6), and will be seen still at the turn of the twenty-first century.<sup>17</sup> But the why and the how still elude us. For description is not theory, and only theory is explanation.

<sup>17</sup> A series of canonical correlations between the set of variables in the late sixties and those of all other time-slices confirmed the obvious as far as the sixty-four major nodes are concerned: namely, that many linear combinations of the variables can be found that relate the sets to an extremely high degree even when the time-span is over forty years.



# 11 The Spatial Organization of Power in the Development of Urban Systems

John Friedmann

## *Power, Exchange, and Spatial Integration*

The study of urban systems in the context of national development is a relatively recent interest. Research has converged on two central questions: what variables account for the growth and development of urban systems? And, how is the growth and development of urban systems related to the more encompassing processes of national development? By 'development of urban systems' I mean the structural growth of urban settlement measured by population and the volume of economic activities. 'National development,' on the other hand, is used here as a shorthand expression for the structural transformation of a national economy to industrialism. Although these questions are clearly not the only ones deserving consideration, they have so far received most of the attention.

The linkages between urban and national development are still inadequately understood. It has nevertheless become clear that their study must employ an explicit spatial framework for analysis. The emergence of modern industrial enclaves within the matrix of an agrarian economy has given rise to dramatic shifts in population and employment and has accelerated urbanization. At the same time, urbanization seems to have been generating its own dynamics, in partial autonomy of the development of modern industry. These complex changes, occurring over the vastness of a national territory, have decisively affected the possibilities of national integration, by demanding new political loyalties, creating new patterns of transportation, giving birth to new social classes and elites, introducing new sets of 'modernizing' values, and differentially affecting the well-being and life chances

*Acknowledgments.* This essay was prepared for the Work Group on Comparative Urbanization of the Social Science Research Council. The present version is based on a preliminary draft, dated January 1972, but has been completely rewritten. I wish to thank John Hanna, Allen Howard, Richard Morse, Francine Rabinovitz, Edward Soja, and Myron Weiner for their generous and critical discussion of the earlier version.

Reprinted from *Development and Change*, vol. 4, no. 3, pp. 12-50, with permission of the Institute of Social Studies, the Hague.

of every member of the population according not only to who he was but also where he lived.

Regarded in this perspective, the study of urban systems has become the study of national development in its spatial dimension. A key question that may, therefore, be put is how the development of an urban system will affect the character and evolution of spatial integration measured by political institutions, transactions, and social justice.<sup>1</sup>

Students of urbanization have tended to explore economic explanations, such as the distribution of natural resources, the location of transport routes, the organization of markets, and economies of scale and agglomeration. With rare exceptions, they have neglected political explanations and, more specifically, explanations given in terms of the spatial distribution of power.<sup>2</sup> The purpose of this paper is to suggest how the analysis of power relations in a national society may contribute to our understanding of the ways in which urban systems evolve.

The concept of power is one of the most elusive in the social sciences. Here, it will refer to the ability of organizational and institutional actors, located in geographic space, to mobilize and allocate resources (manpower, capital, and information) and intentionally to structure the decision-field of others (i.e., to constrain the decisions of others by policies, rules, and commands). Both governmental and private economic power will be considered. Both kinds of power, I will assume, have the capacity to influence the location decisions of firms and households, the quantity, location, and application of resources, and the flow of innovations. By acting on these variables and, in turn, by being acted upon by them, the spatial distribution of power influences the growth and development of urban systems and, at a higher level of synthesis, also the spatial patterns of integration of a national society.

<sup>1</sup> Little agreement exists on what constitutes spatial integration. The term has come into recent usage by geographers who tend to use it in the sense of connectivity and who are likely to measure integration by functional linkages or transactions between places. This usage has much in common with that of Karl Deutsch and his associates. Political scientists have had a more long-standing concern with integration, particularly at the level of international relations. Current research has been brought together in a book edited by Leon N. Lindberg and Stuart A. Scheingold, *Regional Integration. Theory and Research* (Cambridge: Harvard University Press, 1971). For present purposes, the chapters by Ernst B. Haas and Fred M. Hayward are especially useful.

<sup>2</sup> See, for instance, Irving Louis Horowitz, "Electoral Politics, Urbanization, and Social Development in Latin America," *Urban Affairs Quarterly*, II, 3 (March 1967), pp. 3-35, and John Friedmann, *Urbanization, Planning, and National Development* (Beverly Hills, Cal.: Sage Publications, 1973), chapter 5, "Hyperurbanization and National Development in Chile."



Like capital, power refers to a stock of resources rather than to a flow of these resources in use. It will consequently be distributed either *symmetrically* (referring to the capacities of actors that are roughly equal with respect to a common decision area) or *asymmetrically*. The uses of power, on the other hand, involve exchange relations or transactions which may be either *reciprocal* (regarded as bringing roughly equal net-benefits to the actors involved) or *non-reciprocal*. These distinctions allow us to construct a two-by-two matrix of power and exchange in urban systems (Fig. 1).<sup>3</sup> By shifting the argument to a consideration of urban systems, we are abstracting from the particular relations of power and exchange among actors distributed over the whole of a spatially integrated subsystem of society (a city) or an integrated system of cities.<sup>4</sup> The matrix, in fact, is intended to throw into relief the major forms of spatial integration across such systems.

According to this matrix, urban systems in Quadrants 1 and 2 are integrated on a basis of a rough equivalence of power; in Quadrants 3 and 4, they are integrated on a basis of inequality or dependence with respect to the urban system in Quadrant 1. Simple analogies may help to clarify these relationships.

Fig. 1. *A Model of Power and Exchange Relations in Urban Systems*

| Power Relations | Exchange Relations   |  |
|-----------------|--|--|
|                 | Reciprocal   | Non-Reciprocal   |
| Symmetrical     | 1<br>fully integrated urban system:<br>moral authority predominates  | 2<br>competitive urban system integrated on a basis of limited liability: utilitarian power predominates           |
| Asymmetrical    | 3<br>active periphery of urban system integrated on a basis of protective dependency: utilitarian power predominates | 4<br>passive periphery of urban system integrated on a basis of submissive dependency: coercive power predominates |

<sup>3</sup> The theoretical foundation for this matrix is in part derived from Peter M. Blau, *Exchange and Power in Social Life* (London: John Wiley, 1964).

<sup>4</sup> The systems approach to the study of cities was first formalized by Brian J. L. Berry in a justly famous article, "Cities as Systems Within Systems of Cities," reprinted in John Friedmann and William Alonso (eds), *Regional Development and Planning. A Reader* (Cambridge: The MIT Press, 1964), chapter 6. The original article appeared in 1963.

Under 1 (symmetry *cum* reciprocity), relations are as those between friends: neither dominates the other, and the exchange between them will be in balance. Moreover, the rules governing their conduct with respect to each other are accepted as morally right: the costs and benefits of transactions between them are not closely calculated. This relationship is typical of actors within core regions comprising one or several rapidly growing cities that display strong and complexly interwoven patterns of transaction. Where several cities are so related, the statistical form of the urban system will tend to be lognormal. Moreover, the laws and procedural rules under which transactions occur will not generally be open to challenge; their authority will be accepted as morally legitimate.

Under 2 (symmetry *cum* non-reciprocity), relations are as those between the owners of competing business firms: each transaction is separately negotiated in the hope of striking a bargain, so that commitments made in one period are not necessarily considered binding on decisions in subsequent periods. Although each separate transaction may end by being reciprocal, it will be so to only a limited extent; the ultimate intention of each actor is to gain superiority over his competitor. This would be the case of a loose federation of states each having its own integrated urban system, as in Yugoslavia, where the conditions of every inter-system transaction may themselves become the object of intensive bargaining among would-be equals, with the goods offered in exchange serving as the principal counters in negotiation.

Under 3 (asymmetry *cum* reciprocity), relations are as those between superiors and subordinates in bureaucratic organizations: each stands in need of the other, but for quite different reasons. The former require subordinates to accomplish their intentions, but also to rise in general esteem and power, while the latter need the protective benevolence of their superiors and the guarantee of a job. With respect to the organization controlled by their superiors, subordinates have a contractual relationship that may be renegotiated from time to time, but whose legitimacy is generally not at issue. This is the situation typical of many border provinces, such as Magallanes and Tarapacá in Chile which use their exposed position vis-à-vis Argentina, on the one hand, and Peru and Bolivia, on the other (an always threatened shift from 3 to 2), in bargaining for increased autonomy and economic benefits. (The relations of the Commonwealth of Puerto Rico to the United States is a similar instance; here the threat of national independence serves to strengthen the bargaining position of the Commonwealth.) Active pe-

ripheries are typically striving to build up one or more growth centers as core regions subordinate to the urban system in Quadrant 1. They do so in the hope – however much in the future – of ultimately being absorbed into the fully integrated core region itself.

Finally, under 4 (asymmetry *cum* non-reciprocity), relations are as those between master and slave: the master dominates his slave who, at least outwardly, gives evidence of properly submissive behavior but whose labors on behalf of his master are poorly rewarded. Occasional rebellion on the slave's part may invoke the full repressive power of the master. This is the case of economically backward regions under a regime of internal colonization (such as Bangladesh before independence) which have few cities, and whose domination by the core region in Quadrant 1 gives rise to an urban system having pronounced primacy characteristics. The latent capacity for rebellion by the passive periphery may induce the dominant interests in Quadrant 1 to invest heavily in the region and so to shift it eventually to Quadrant 3. Indeed, such measures may occasionally be taken for purely ideological reasons. On the other hand, the failure to invoke coercive power may result in little more than spreading anarchy without compensating economic benefit. This may be illustrated with reference to the recent economic collapse of the agricultural system in Chile's southern provinces or the continued agitation under Salvador Allende's Popular Unity Government of extreme left-wing revolutionary groups centered in the city of Concepción. Passive peripheries no longer fully dominated by the core in Quadrant 1 may eventually come within the area of influence of the competing system in 2. They have little strength of their own to resist such advances, and their original oppressor may be equally incapacitated.

In the following four sections, some of these relations of power and exchange in urban systems will be further analyzed. First, I shall try to show how the spatial distribution of governmental power influences the location decisions of entrepreneurs during the early phases of industrialization and how the growing interpenetration of governmental and private economic institutions channels the subsequent location decisions of individuals and households to locations of central power in excess of objective opportunities for productive employment. The resulting polarized pattern of urbanization tends to be self-perpetuating, whereas the eventual decentralization of productive activities into the passive periphery of major core regions tends to leave essential relations of power virtually unchanged.

The second example relates to the diffusion of innovations through

the urban system. I will be concerned only with entrepreneurial innovations whose successful adoption translates into a relative increase in economic power to exploit specific resources in the environment. The diffusion of innovations will be considered in both space and time. The spatial diffusion of entrepreneurial innovations tends to be hierarchical, leading to a steadily increasing concentration of power in the largest cities of the urban hierarchy, while the rate of diffusion, at least initially, gives special advantage to early over late adopters. The resulting growth pattern of cities tends to be allometric, implying invariant ratios in the rates of growth among individual urban units. Passive peripheries are thus 'condemned' to a quasi-permanent condition of submissive dependency, though the active portions of the periphery may be able successfully to negotiate for growing autonomy in development decisions.

The third illustration concerns primarily the conflict patterns between competing economic and political elites, where the former are ethnically and/or culturally distinct from the latter and have primarily an urban base, while the latter's base of power tends to be in rural areas. Several options for resolving conflicting interests will be discussed, including cooptation, accommodation, open hostility, the creation of regional protectorates, and federative solutions, each of which will have different outcomes for the development of the relevant urban systems.

In the final section, a case study of dependency relations in Chile will be presented. Various forms of dependency will be discussed, together with their consequences for the development of urban systems in this small South American country.

No effort will be made to synthesize these four approaches to the study of power relations in urban systems. The paradigm presented in this section is intended to serve primarily as a source of hypotheses for testing in empirical settings. For this reason, too, I shall make no effort to append a section on policy options. At this stage in our knowledge, such an exercise would be gratuitous. The only firm conclusion we may draw is that the process of national development and spatial integration is an eminently political one, involving fundamental relations of power and exchange and the resolution of resulting conflicts. Planning which fails to recognize this basic truth and proceeds as though the spatial allocation of resources were merely an exercise in applied rationality is bound to be disappointing in its results.<sup>5</sup>

<sup>5</sup> The scientific bases for prescriptive policies of urban development are still weak. But even if they were stronger, it is unlikely that they would provide unambiguous conclusions for optimal courses of action. A brilliant review of the

*Economic Location and the Spatial Distribution of Power*

Economic location theory has traditionally addressed the question of how the location decisions of individual firms are affected by spatial variations in the costs of production and distribution. This emphasis reflects in part the observations of location theorists in industrially mature economies. In countries of incipient or early industrialization, however, non-economic influences appear to weigh more heavily in location decisions than considerations of relative cost. In these countries, *the choice of a location tends to be strongly influenced by a desire of management to gain direct access to the relevant centers of governmental power.*

In the following, I shall assume an industrializing country of moderate size whose government is unitary and whose population is culturally homogeneous. Subsequently, I shall relax this assumption, but for now it will serve as a necessary constraint. In such a country, economic enterprise is exceedingly dependent on the central bureaucracy and the corridors of legislative power. Licenses to import machinery must be secured; special subsidies and other favors are sought; a complex system of legislation pertaining to the conduct of business must be learned; and contributions of public capital and credit are expected. At the same time, economic interrelationships are relatively weak: an inter-industry matrix would show many empty boxes.

In themselves, these conditions would not prescribe a central location. They are reinforced by additional considerations that make the creation or survival of new enterprise in provincial districts highly improbable. Among them are (a) a still rudimentary system of transport and communication, (b) the great importance attached to personal, face-to-face relations in the conduct of business, (c) a high degree of bureaucratic centralism, and (d) a superior infrastructure of economic and social facilities in the national capital, itself a reflection and symbol of accumulated (and steadily accumulating) power.

The resulting symbiosis between economic and governmental organizations creates a situation that consistently favors the nation's capital in subsequent business locations, though economic reasons, such as access to markets, undoubtedly contribute. Politicians, bureaucrats and businessmen mingle in exclusive social clubs and the city's top restaur-

---

current state of knowledge in urban systems analysis, from a perspective of public policy, is Harry W. Richardson's, "Optimality in City Size, Systems of Cities in Urban Policy: A Sceptic's View" (Centre for Research in the Social Science, University of Kent at Canterbury, Reprint Series No. 18 (1), 1972).

rants, send their children to private schools (or the national university), and form tight social networks of their own. From this central location, an essentially passive periphery is organized into administrative and market areas following the principal routes of transport. Capital resources and surplus agricultural labor are withdrawn from these areas at an accelerating pace, adding to the reservoir of economic power in the center. In consequence, the urban pattern changes from one of low-level equilibrium (many small, equally sized urban places) to one of growing primacy.<sup>6</sup>

With continuing development, however, certain changes in this spatial pattern may occur. Growing markets, the discovery of new natural resources, and a gradually improving system of transport and communications may render middlesized cities in the periphery increasingly attractive as possible business locations, a tendency that may be actively encouraged by explicit governmental policies for regional development. These changing circumstances, together with the growing organizational complexity of enterprise, make possible the physical separation of management from production units. With their vital decision functions thus removed, production units are released to locate according to economic criteria, while management components continue to be drawn to the center of governmental power. Even so, it is generally provincial administrative centers that are favored in the location of production units to facilitate the symbiotic decision process that governs the economic life of the nation.

Empirical evidence for this evolving pattern comes from a variety of country settings. For Latin America, the historian Richard M. Morse is quite emphatic. He writes:<sup>7</sup>

<sup>6</sup> Much controversy has raged over the issue of whether the size distributions of cities is anything but an empirical curiosity. A great deal has been written specifically about the form of rank-size distributions and whether these are in any way related to conditions of economic development and integration. In a recent piece, Brian J. L. Berry, who has been in the center of this controversy, has revised his earlier view that the evidence for a clear-cut relationship is inconclusive. Basing his argument on time-series data for change in size distributions (whereas earlier analysis had been restricted to comparative statics), he now maintains that urban systems typically evolve from a low level equilibrium distribution (many small, equally sized urban places) via urban primacy to a high-level equilibrium characterized by a lognormal distribution of city-sizes. See his "City Size and Economic Development: Conceptual Synthesis and Policy Problems, with Special Reference to South and Southeast Asia," in Leo Jacobson and Ved Prakash (eds), *Urbanization and National Development* (Beverly Hills, Cal.: Sage Publications, 1971), ch. 5.

<sup>7</sup> Richard M. Morse, "Planning, History, Politics," in John Miller and Ralph A. Gakenheimer (eds), *Latin American Urban Policies and the Social Sciences*

In Latin America, it seems important that a city be a patrimonial center if it is to serve as a growth pole for economic development. Brasília is already the classic case for a modern frontier zone. Or, if a capital is not actually transferred to a frontier, the central power may spin off an outlying city under its direct support and tutelage, as in the case of Ciudad Guayana. Without denying the regional economic and ecological justifications for this city, it is probably accurate to say that its ultimate legitimation derives from a process of patrimonial schizogenesis. Or again, if planners speak of decentralizing economic functions from a central corridor not to a frontier but to existing peripheral cities, it is usually implied that provincial capitals will be the beneficiaries. Thus it is no accident that the flourishing second-echelon growth centers (Monterrey, Guadalajara, Cali, Medellín, Córdoba, Porto Alegre, Curitiba) are so frequently regional political capitals. When this is not the case, as with Chimbote, Peru, the city may face enormous obstacles in developing urban infrastructure for economic activity because of its weak political leverage.

The second example refers to the Soviet Union and is reported by Chauncey Harris.<sup>8</sup>

The importance of administrative and related functions is expressed in the relatively rapid growth of *oblast* centers. In about 60 percent of the *oblasts*, the center grew more rapidly than other urban units within their boundaries.

The third example comes from J. Barry Riddell's study of the spatial dynamics of modernization in Sierra Leone.<sup>9</sup>

Thus it is evident that the process of modernization, as summarized by the component analysis, is dominated and directed by the network and the [urban-administrative] hierarchy, which together define the spatial fabric of the country.

The fourth example stems from Brazil, a country that has moved considerably beyond the first thresholds of industrialization. The concentration of modern business enterprise in Brazil was initially confined to the two principal centers of economic power: Rio de Janeiro and São Paulo. By the time the political capital of the nation was shifted to Brasília in the latter part of the 1960s, industrialization had already established a powerful base in these two cities and, to a much smaller

(Beverly Hills, Cal.: Sage Publications, 1971), p. 194.

<sup>8</sup> Chauncey D. Harris, "Urbanization and Population Growth in the Soviet Union, 1959-1970," *Ekistics*, 32, 192 (November 1971), p. 360.

<sup>9</sup> J. Barry Riddell, *The Spatial Dynamics of Modernization of Sierra Leone* (Evanston: Northwestern University Press, 1970), pp. 90-93.



extent, in several of the more important state capitals (Belo Horizonte, Pôrto Alegre). Because economic power had now become more important than political power, the physical move of governmental functions to Brasília did not entail a similarly massive shift of corporate headquarters to the planalto of Goiás, though it did much to stimulate road building activity and cattle raising in the interior. By the same token, intensive government efforts to industrialize the traditionally backward regions in the North and Northeast of the country accomplished primarily the move of production units to these regions but failed to attract units of corporate management. With management remaining in the older centers, and attracting related business services, the 'decapitated' production units in the periphery found themselves dependent on extra-territorial decisions. Business profits, in particular, were transferred to the 'center' for reallocation.

The evidence for the pattern described is impressive. While political and economic decision-making power remain concentrated in the national capital, subsidiary growth centers spring up on the periphery, frequently paralleling the urban-administrative hierarchy. This process tends to induce a gradual filling out of the rank-size distribution of cities by encouraging the growth of intermediate urban centers. As a result, certain portions of the passive periphery may be activated sufficiently to bargain with central authorities for greater autonomy (e.g., the Northeast of Brazil). To the extent they are successful, the dependency relations of the remaining periphery may increasingly come to focus on these subsidiary, provincial centers.<sup>10</sup>

If we carry the analysis still further to include advanced industrial and post-industrial societies, the earlier pattern, though in a highly attenuated form, may still be discerned.<sup>11</sup> By this time, the extreme de-

<sup>10</sup> Success in bargaining may depend on the strength of a number of variables, including the size of region, the ethnic/cultural composition of the region's population compared to that of core elites, the relative location of the region in terms of distance from the core and proximity to international frontiers, the unitary or federal structure of the government, and political finesse.

<sup>11</sup> The most impressive evidence comes from a Swedish study by Gunnar Törnqvist, *Contact Systems and Regional Development* (Lund Studies in Geography, Ser. B, Human Geography No. 35, Lund: C. W. K. Gleerup, 1970). For the United States, a statistical study of non-production personnel in manufacturing similarly suggests that locational separation between managerial and production functions exists, and that the former tend to be found in the larger, more rapidly growing metropolitan areas. See Esther Emiko Uyebara, *Production and Non-production Employment in Manufacturing. A Comparative Analysis of Metropolitan Areas* (Master's thesis, School of Architecture and Urban Planning, UCLA, 1972).



pendency of business on governmental power may have waned relative to the rapidly growing requirements for inter-industry contacts. Both market and supply areas will have become more diffused, and the transport and communications system will have made the relevant economic space more accessible from a larger number of central locations. Parts of the formerly active periphery may by now be effectively integrated into the principal core areas of the nation. Despite these new developments, however, certain nodal cities may still stand out as 'control centers', experiencing rapid growth, even though the initial close linkage between centers of governmental power and business location will have been lessened.<sup>12</sup> The urban system will now tend toward a lognormal form in the distribution of its centers, and the passive periphery will be reduced to vestigial proportions.<sup>13</sup>

The foregoing description of the evolution of a spatial system is, of course, idealized to some extent. Small countries with only one or two major cities, very large countries such as the USSR, China, and India with a long-standing tradition of urbanism, countries with a federal structure of government, and countries with a culturally heterogeneous and regionalized population may follow a different sequence of events. In actively federal systems, for example, central power will, to some degree, be shared so that several governmental centers may simultaneously compete for industry (Yugoslavia). By the same token, regions having politically powerful minorities may gain certain privileges, such as greater decision autonomy, sooner than would be predicted by the model. In these situations, the 'idealized' spatial pattern may be distorted for the nation as a whole, though the pattern is likely to be replicated at the regional level.<sup>14</sup> Furthermore, once they are established, spatial

<sup>12</sup> Empirical evidence supporting a concept of nodal city is found in Thomas M. Stanback, Jr., and Richard V. Knight, *The Metropolitan Economy* (New York: Columbia University Press, 1970), *passim*.

<sup>13</sup> I am assuming a strong connection – still to be demonstrated mathematically – between Brian Berry's model of the evolution of city size distributions and Jeffrey Williamson's model of the evolution of regional inequalities of income. Williamson argues that regional income diverges from the mean during the early stages of economic development (analogous to the emergence of urban primacy under the first impacts of development) but subsequently, if gradually, converges as development proceeds (analogous to Berry's approximation to a lognormal distribution of city sizes). An explicit spatial mapping of these two processes has not yet been accomplished, however. For Berry, see footnote 6. Williamson's model was published as "Regional Inequality and the Process of National Development," *Economic Development and Cultural Change*, Part II (July 1965), pp. 3-45.

<sup>14</sup> For India, Brian Berry has found four core regions of approximately equal influence and through which India's space economy appears to be organized.

patterns of urbanization tend to perpetuate themselves, casting a long shadow into the future.<sup>15</sup> The initial distribution of governmental power within a country will therefore tend to guide the subsequent evolution of the space economy.

*The Spatial Diffusion of Innovations  
in the Development of Urban Systems*

Studies of the spatial diffusion of innovations have only recently begun to turn from an exclusive concern with questions relating to geographic theory to broader issues of socio-economic development. These newer studies strongly suggest the possibility of interpreting the spatial dimensions of all facets of development, including urbanization, from a perspective of innovation diffusion. Although a parsimonious theory of the observed behavior is still some time away, its major contours are beginning to be seen.<sup>16</sup> An important link in such a theory is the relation of spatial diffusion to the distribution of economic power.

They are based, respectively, on Bombay, Delhi, Calcutta, and Madras. See Fig. 3 in Berry, "City Size and Economic Development," *op. cit.*, p. 121.

<sup>15</sup> The strongest case, both theoretically and empirically, for the stability of the spatial and size distributions of urban systems comes from J. R. Lasuén, "Multi-Regional Economic Development. An Open System Approach," in Torsten Hägerstrand and Antoni R. Kuklinski (eds), *Information Systems for Regional Development* (Lund Studies in Geography, Ser. B. Human Geography, No. 37, Lund: C. W. K. Gleerup, 1971), pp. 169-211. His findings are supported for the People's Republic of China by Yuan-Li Wu, *The Spatial Economy of Communist China. A Study of Industrial Location and Transportation* (New York: Frederick A. Praeger, 1967).

<sup>16</sup> The starting point for the study of the spatial diffusion of innovations is Torsten Hägerstrand, *Innovation Diffusion as a Spatial Process* (Chicago: The University of Chicago Press, 1967. Original in Swedish, 1953). A comprehensive annotated bibliography of spatial innovation diffusion studies through 1968 has been compiled by Lawrence A. Brown, *Diffusion Processes and Location. A Conceptual Framework and Bibliography* (Bibliography Series no. 4, Regional Science Research Institute, Philadelphia, 1968). The relation of spatial diffusion processes to economic development is worked out by Allan R. Pred, *Behavior and Location. Foundations for a Geographic and Dynamic Location Theory*, Part II (Lund Studies in Geography, Ser. B. Human Geography, No. 28, Lund: C. W. K. Gleerup, 1969), chapter 4, and by John Friedmann, "A Generalized Theory of Polarized Development," in Niles Hansen (ed), *Growth Centers in Regional Economic Development* (New York: The Free Press, 1972). Lasuén's study (see footnote 15) is also relevant here, as is Edgar S. Dunn's pathbreaking study, *Economic and Social Development. A Process of Social Learning* (Baltimore: The Johns Hopkins Press, 1971). The basic reference for innovation diffusion studies generally is Everett M. Rogers, *Communication of Innovations. A Cross-Cultural Approach* (Second Edition, New York: The Free Press, 1971).

The basic thesis – to be elaborated in the following pages – may be briefly stated at the outset. The adoption of innovations, and particularly of entrepreneurial innovations (see below), *translates directly into an increase of effective power by the adopting unit over portions of its environment*. The firm adopting a corporate structure may push more traditionally organized competitors out of business; or the manufacturer introducing a piece of new machinery may improve the quality of his product (or lower his costs), capturing a larger share of the market. The cumulation of entrepreneurial innovations in a given city – the city being conceived as a spatially integrated subsystem of society – will therefore lead not only to its accelerated economic and demographic growth, but also to the consolidation of its hierarchical control over that portion of the urban system that has failed to adopt this particular set of innovations. Such a concentration of innovations in cities that have a high propensity for further innovation, produces the wellknown phenomenon of core regions that extend their control over the dependent peripheries of the country and, in some cases, abroad. The basic relations in the spatial distribution of economic power are thus seen to be an immediate outcome of the diffusion of innovations. Only a concerted governmental effort to establish conditions favorable to accelerated innovation at selected points in the periphery is likely to produce a marked reorganization of a growth pattern that, under normal conditions, displays remarkable stability. This stability, it turns out, is itself the result of innovation diffusion processes.

The extensive literature on innovation diffusion is generally deficient in that it fails to distinguish among broad categories of innovations. Basing a criterion of classification on structural form, for example, product, cultural, technical, and organizational innovations may be distinguished. Alternatively, a classification based on salient characteristics of the adopting unit suggests a grouping into *consumer* and *entrepreneurial* innovations. These two systems of classification may be combined as follows:

- a. consumer innovations: product and cultural (related primarily to the *demand* side of economic transactions)
- b. entrepreneurial innovations: technical and organizational (related primarily to the *supply* side of economic transactions)

This simplified system has the merit of facilitating the integration of spatial diffusion studies with economic theory. If, as seems probable, consumer innovations diffuse more rapidly and over wider areas than

entrepreneurial innovations, a ready explanation for the spatial dynamics of the development process would seem to be at hand. Pressures for development arise from the side of demand (itself the result of prior diffusion processes) and occasion vast population migrations to the principal centers of entrepreneurial innovation where, it is hoped, these demands can be satisfied more expeditiously. This hope, of course, is usually disappointed. Only an explicit policy to contain the diffusion of consumer innovations (as in socialist economies) is able to reduce the level of demand sufficiently to permit the carrying out of a broader policy directed at a sustained and long-term increase in the supply bases of the economy.

In the remainder of this section, I shall refer exclusively to entrepreneurial innovations. Unfortunately, the empirical evidence for this type of innovation is slim compared to that available for consumer innovations. In the absence of sufficient studies, I am constrained to put forward a series of plausible but largely untested propositions that may hopefully serve as a basis for future comparative research.

*Proposition 1.* The spatial diffusion of entrepreneurial innovations follows the paths of exchange relations among cities. However, regardless of where in a system of cities an innovation enters, it will soon be captured by the largest city or cities in the system.

Evidence for this proposition, modified to include all growth-inducing innovations, has been brought together by Allan R. Pred who has also given this proposition an elegant mathematical formulation.<sup>17</sup> According to Pred, the strict hierarchical diffusion model, according to which innovations proceed in orderly progression down the urban size hierarchy of cities must be abandoned. This is true for the general case. In the ideal-typical developing country (moderate size, unitary government, culturally homogeneous population), however, the entry points for most innovations tend to be the largest, most cosmopolitan cities, such as the national capital or major port cities, and inter-urban contact networks tend to be hierarchical with respect to these cities; the number of non-hierarchical linkages are few. Where this is the case, the diffusion process will tend to be hierarchical even while obeying the general law

<sup>17</sup> Allan R. Pred, "Large-City Interdependence and the Preelectronic Diffusion of Innovations in the U.S.," *Geographical Analysis*, 3 (1971), pp. 165-81. The more complete formulation of this model is presented in an unpublished manuscript by the same author, "Interurban Information Circulation, Organizations and the Development Process of Systems of Cities" (Department of Geography, University of California, Berkeley, 1972).

governing spatial diffusion processes formulated by Pred. Moreover, some innovations, because of the intrinsic uniqueness or scale relative to the size of the national economy, never diffuse beyond the points of initial adoption and may, therefore, be regarded as *national* innovations (e.g., a stock exchange or oil refinery). For similar reasons, other innovations may be limited to only one per region (e.g., a hydro-electric installation) or one per city of a certain size (e.g., a municipal water works) and may thus be called *regional* and *urban* innovations respectively.

The accumulation of national innovations in only one city will give that city a preeminent role in directing the country's economic affairs. By analogy, the same will happen at lower hierarchical levels with cities that rapidly accumulate major regional and/or urban innovations. A hierarchy of urban centers exercising control over both national and regional economies is thus established. For reasons already stated, economic control centers will frequently coincide with centers of governmental power, so that the two hierarchies – economic and administrative – may eventually be joined.

*Proposition II-A.* Especially during the starting-up phases of development, increased economic advantage accrues to the early adopters of innovations. To the extent that this 'initial advantage' is translated into vigorous urban growth, cities receiving the largest number of early innovations will tend to experience more rapid growth than cities adopting the same innovations later in time. With continuing development, however, the period required for a complete cycle of diffusion tends to diminish, so that smaller cities will increasingly come to share in exploiting the innovations in question, accelerating their own growth. The time sequence of innovation diffusion will nevertheless continue to be an important influence in the spatial distribution of economic power, since the rate at which innovations enter the urban system, relative to their downward diffusion, will tend to increase in the upper reaches of the hierarchy (see III below).

*Proposition II-B.* Especially during the early phases of development, the rate of diffusion will tend to be faster for centers in proximity of the initial points of adoption than for centers of equivalent rank located at greater distance from these points.

Rogers distinguishes between (a) innovators, (b) early adopters, (c) early majority, (d) late majority, and (e) laggards whose distribution in



time tends to follow the shape of an S-curve.<sup>18</sup> The slope of this curve will be different for each innovation, but will generally tend to rise over the period of development reflecting, among other things, improved transport and communication linkages, larger organizational scale, wider contact networks, and the accumulation of earlier innovations. Innovators and early adopters (counted as individual cities) will thus enjoy a quasi-monopolistic position in exploiting innovations before these innovations spread to other adopting units.

The diffusion process is governed by underlying patterns of information-exchange, especially of face-to-face communication. The spatial pattern of information exchange, however, is subject to declining intensity with distance, or distance decay. During the early period of a country's development, the distance decay curve tends to be rather steep (localism predominates), but eventually it tends to flatten out; communication processes become less constrained by distance, and other variables acquire greater salience. Centers located near points of initial adoption are therefore likely, *ceteris paribus*, to receive innovations earlier than centers of equivalent rank located at greater distance. This will tend to forge strong complementary links among adjacent centers, a process conducive to the formation of multi-centered core regions clustered around the largest, most innovative, and cosmopolitan cities of the country.

*Proposition III.* The probability of entrepreneurial innovations is an increasing function of city size. The larger the city in the size of its effective population, the greater will be the probability of innovation.<sup>19</sup>

This hypothesis, which underlies much of the preceding discussion, in turn depends on a number of intervening variables. In the following list, each variable is assumed to be an increasing function of city size.

1. *The demand for innovations.* Large cities have a greater need than small cities for innovations in helping solve new problems resulting from accelerated growth, growing population densities, increased specialization, and greater structural complexity. At the same time, organizations located in large cities tend to have a greater capacity for searching out potentially useful innovations than organizations in smaller centers.
2. *The financial, technical, and organizational resources for innovation.* Organizations in large cities tend to have greater access to and are

<sup>18</sup> Everett M. Rogers, *op. cit.*, p. 27.

<sup>19</sup> To make this proposition true, population size must be standardized for education and possibly also for income.

- able to mobilize resources for innovation more effectively than organizations in smaller cities.
3. *The propensity to innovate.* Innovative talent tends to move up the urban hierarchy within a country and down the hierarchy from external core regions to the largest, most cosmopolitan cities. The frequency of entrepreneurial skills in the population is therefore greater in large than in small cities.
  4. *Cultural receptivity to innovations.* For many contemporary innovations, the requirements for receptivity – cultural, educational, linguistic, and technical – can be formidable. Persons having the skills which enable them to perceive the advantages of an innovation and also have the technical knowledge to carry them through tend to be more prevalent in large than in small cities.
  5. *The stock of information available to potential innovators.* There is some evidence that the stock of available information to individuals and organizations increases exponentially with city size.<sup>20</sup> Large cities are information-saturated environments. The density of information is positively correlated with the probability of information.
  6. *The range of contact networks.* The presence of information and a generalized receptivity are not in themselves sufficient for effective communication. Potential innovators must be tied into contact networks through which the relevant information is passed on. These networks tend to be more extensive, and there is probably greater redundancy of information, for organizations located in large than in small cities.
  7. *Structural compatibility of innovations.* Before they can be implemented, many innovations require complementary innovations, such as supporting services. These are more likely to be present in large than in small cities. In addition, repeated innovation experience creates attitudes and expectations favorable to further innovation. Innovation in large cities tends to become an institutionalized process.
  8. *Employment multipliers of innovation.* This relationship rests on the idea that economic specialization, which tends to rise with increasing city size, implies higher employment multipliers from the adoption of an innovation. Multipliers are generated by the linkages of an innovation with supporting sectors. Large cities may also help support innovations in smaller cities, thus 'capturing' a part of the employment multiplier of such cities.
  9. *Economic thresholds for innovation.* Innovations become economically feasible only at certain threshold sizes of total income and population. By definition, these thresholds increase with city size. In addition, there is some evidence that the threshold values for innovations have themselves been increasing over the period of industrialization.

<sup>20</sup> Fascinating data in support of this hypothesis have been brought together by Toshio Sanuki, "The City in Informational Society," *Area Development in Japan*, 3 (1970), pp. 9-23. Sanuki's study is frustrating, however, because he does not reveal the basis of his calculations.

*Proposition IV.* Over the course of development, the character of innovations changes in the direction of growing economic size, rising costs but also higher productivity, increasing specialization, and increasing technical complexity.

In the terminology adopted here, this proposition implies a relative increase in the number of national and regional innovations. Alternatively, we may say that the urban threshold sizes for innovation tend to rise with development. Entrepreneurial innovations will, therefore, tend to diffuse over progressively shorter hierarchical distances, assisting the growing polarization of development and leaving lower-order centers in a steadily worsening position, as both population and capital flow *up* the urban hierarchy in search of greater opportunity. These urban patterns will be reflected in growing regional differences in the levels of per capita income and other indices of socioeconomic development.

*Proposition V.* The adoption of an innovation in period I increases the probability of further innovation by the adopting unit in period II.

Innovation may be understood as part of a learning process in which prior success predisposes an actor to further innovation at an accelerating rate.<sup>21</sup> Clearly, there are upper limits to the rate of innovation, the capacity for continuous innovation rising progressively from individuals, to organizations, to society. But the existence of such limits does not deny the positive influence that learning has on the growth curve of innovation, as search behavior improves, and the entire process of introducing innovations into an existing system becomes routinized.

Rogers identifies the following characteristics of early adopters:<sup>22</sup>

Earlier adopters... have greater empathy, less dogmatism, greater ability to deal with abstractions, greater rationality, and more favorable attitudes toward change, risk, education, and science. They are less fatalistic and have higher achievement motivation scores and higher aspirations for their children. Earlier adopters have more social participation, are more highly integrated with the system, are more cosmopolite, have more change agent contact, have more exposure to both mass media and interpersonal channels, seek information more, have higher knowledge of innovations, and have more opinion leadership.

These characteristics are not inborn traits, however. They can be learned

<sup>21</sup> Everett M. Rogers, *op. cit.*, p. 178. See also Edgar S. Dunn, Jr., *op. cit.*, and Allan R. Pred, *op. cit.*

<sup>22</sup> Everett M. Rogers, *op. cit.*, p. 196.



in the course of successful innovation. In the long run, an entire society may learn to be innovative, but initially the rapid learners will be found predominantly among the populations of large, cosmopolitan cities where innovations tend to be initially introduced. As a result, the rate of innovation in these cities is likely to be higher than the rate of diffusion of these innovations to other parts of the urban system, increasing the differences among centers in regular hierarchical sequence.

*Concluding Comments.* Except for Proposition 1, the innovation diffusion process described above follows closely the theoretical model evolved by J. R. Lasuén. His conclusions are worth quoting in full.<sup>23</sup> (In the following quotation the phrase 'urban system' may be substituted for 'system of regions'.)

In our view, the system of regions grows and develops in a stable hierarchical order due to the factors maintaining the stability of the geographical diffusion patterns (stability in the functional diffusion patterns and rigidity in the firms' locations).

Within each innovation set, the regions grow at differential rates (keeping the stable hierarchical order) due to the effect of the factors which control the feasibility of adoptions in the different regions (diffusion times, market sizes) in interaction with the values of the main characteristics of the innovation (scale of operations, adoption times).

Over time, the values of the innovation set characteristics change (scales of operations increase; adoption time shortens). This causes further differentiation of regional growth rates.

Consequently within every innovation set, regions grow stably hierarchized and allometrically. Over several innovation sets they also grow hierarchically stable and allometrically, but the allometries for every set have different values (normally of successively rising slopes).

In other words, the diffusion of innovations is such that size hierarchies of cities are maintained over successive cycles of diffusion, but the specific economic values captured over the entire system tend to rise in geometric progression from low-ranking to high-ranking cities. Top cities in the hierarchy will consequently adopt more innovations per unit of time than other cities in the system, spinning off older, less efficient innovations to the periphery. This process accounts for the frequently observed sliding scale of diminishing modernity and power as

<sup>23</sup> J. R. Lasuén, *op. cit.*, p. 191.

one descends the urban hierarchy. Towards the upper end of the hierarchy we find a preponderance of metropolitan types with far-flung contact networks while, lower down, narrowly circumscribed fields of interaction and limited horizons of aspiration, knowledge, and opportunity are more prevalent. This pattern corresponds to Allan Pred's large-city-focused model of urban systems growth in which a small but relatively stable set of large cities (or core regions) exerts decisive influence over the growth patterns of a larger set of lower-ranking peripheral regions. The resulting socio-economic indices have been carefully charted by Brian J. L. Berry.<sup>24</sup> The economic landscape of a country is cleft by huge troughs of economic backwardness that divide occasional peaks and ridges of high growth and material wellbeing.

This 'normal' patterning of urban growth can be altered only by changing the distribution of intervening variables and attracting production units into the periphery that are innovation-prone and likely to produce large employment as growth pole policies.<sup>25</sup> The selective activation of 'growth poles' in the periphery will, of course, merely replicate the national pattern of innovation diffusion on a regional scale. At this point the question remains unresolved over how many levels in the urban hierarchy this process of activating growth poles may be extended, and when the normal filtering processes of innovation must be allowed, for lack of suitable controls, to operate without policy intervention.

#### *Interregional Patterns of Conflict and Accommodation*

Innovative entrepreneurial elites in urban areas are frequently found among foreign or national ethnic (or cultural) minorities. Although the entrepreneurial role of foreign 'colonial' elites is generally recognized, national minorities which have gained control over significant portions of the modern economic sector are equally important. The Jews in Western Europe were an early instance of such an elite. In the newly industrializing countries, the Chinese in Malaysia and Indonesia,<sup>26</sup> the

<sup>24</sup> Brian J. L. Berry and Elaine Neils, "Location, Size, and Shape of Cities as Influenced by Environmental Factors: the Urban Environment Writ Large," in Harvey S. Perloff (ed), *The Quality of the Urban Environment. Resources for the Future* (Baltimore: The Johns Hopkins Press, 1969), ch. 8.

<sup>25</sup> Antoni Kuklinski and Ricardo Petrella (eds), *Growth Poles and Regional Policies* (The Hague: Mouton, 1972); and Antoni Kuklinski (ed), *Growth Poles and Growth Centers in Regional Planning* (The Hague: Mouton, 1972).

<sup>26</sup> T. G. McGee, "Têtes de ponts et enclaves. Le problème urbain et le processus d'urbanisation dans l'Asie du Sud-Est depuis 1945," *Tiers Monde*, XII, 45

Ibo in Nigeria,<sup>27</sup> the Antioqueños in Colombia,<sup>28</sup> and the Arabs, Italians, Germans, and Jews in Latin American countries<sup>29</sup> are frequently cited examples of urban innovative elites. (Other ethnic minorities whose entrepreneurial roles might be studied include English Canadians in Quebec, Arabs in Zanzibar, Indians in Burma, East Africa, Trinidad, and the Guayanas, Greeks in Egypt, Slovenes in Yugoslavia, and French settlers in Algeria.)

In nearly every instance, urban ethnic minorities operate in a political environment that is initially controlled by an agrarian-based governing elite whose members belong to a different cultural, ethnic, or religious group. This situation is dramatically illustrated by data on East African cities. According to William and Judith Hanna, "In Kenya, 3 out of 100 residents are non-Africans, whereas in Nairobi the figure is 41 out of 100. Similarly, Uganda's population is just over 1 percent non-African, but for Kampala the percentage is 49. Comparable situations are found in Tanzania and Zambia."<sup>30</sup> And they continue: "With independence, some Africans moved to the top and, as a corollary, Asians and Middle Easterners have been left in a somewhat ambiguous position: subordinate to the new African elite, but on some measures superordinate to the African rank-and-file. The ambiguity arises because racial boundaries prevent Asians and Middle Easterners from entering a unilinear status hierarchy."<sup>31</sup> Many of these non-Africans were, in fact, born on the continent, but remain alien to the indigenous cultures.

Where innovative entrepreneurial elites are excluded from political power, a profound disjunction occurs between rural and urban development. Cities which have the largest concentration of innovative ethnic (cultural) minorities will experience the most rapid growth, while 'na-

(1971), pp. 115-14; Clifford Geertz, *Peddlers and Princes. Social Change and Economic Modernization in Two Indonesian Towns* (Chicago: Chicago University Press, 1963); and Allen E. Goodman, "The Political Implications of Urban Development in Southeast Asia: The 'Fragment' Hypothesis," *Economic Development and Cultural Change*, 20, 1 (Oct. 1971), pp. 117-30.

<sup>27</sup> Robert A. Levine, *Dreams and Deeds: Achievement Motivation in Nigeria* (Chicago: University of Chicago Press, 1966).

<sup>28</sup> Everett E. Hagen, *On the Theory of Social Change* (Homewood, Ill.: The Dorsey Press, 1962), ch. 15.

<sup>29</sup> Seymour Martin Lipset, "Values and Entrepreneurship in the Americans," chapter 3 in *Revolution and Counter-Revolution* (Rev. ed., New York: Anchor Books, 1970).

<sup>30</sup> William John Hanna and Judith Lynn Hanna, *Urban Dynamics in Black Africa* (Chicago: Aldine-Atherton, 1971), p. 109.

<sup>31</sup> *Ibid.*, p. 111.

tive' centers, tied to the rural economy in the periphery, are likely to stagnate. Under conditions of rural/urban disjunction or *economic dualism*,<sup>32</sup> urban-generated surpluses tend not to be used for developing the rural sector (which contains a majority of the total population), but are accumulated, in part to build up the modern commercial-industrial complex at the core and, in part, to be expatriated to the home country of the intruding elite. By the same token, innovations will be contained largely within the core because contact networks and investment resources will also tend to be ethnically (and culturally) controlled. As a result, the remainder of the country will supply the urban core with food, raw materials and labor and, in turn, provide a market outlet for certain core region products.

In situations of this sort, relations between innovative (urban-economic) and traditionalist (rural governing) elites will be variously characterized by patterns of *coöptation*, *accommodation*, and *open hostility*.

Under *coöptation*, the governing elite is placed in a client relation to the entrepreneurial elite. This is typically the case where the latter is of foreign extraction and unassimilated to the national society (Americans in Venezuela under Pérez Jiménez, Japanese in occupied Korea, English in colonial Nigeria and Ghana, Russians in the former Baltic countries, Germans in Norway and France during World War II). For Spanish-speaking Latin America, it has been argued that foreign dependency and coöptation of national elites accounts for the extreme concentration of economic and political power in the national capital regions of countries such as Venezuela, Peru, Bolivia, Ecuador, and Chile.<sup>33</sup> Although this contention remains to be demonstrated, it is claimed that a more integrated form of spatial development will be achieved only if the governing elites regain a substantial measure of autonomy with respect to foreign entrepreneurial elites.<sup>34</sup> In South America, these claims have been advanced primarily by intellectuals, equally hostile to foreign and

<sup>32</sup> For an excellent recent discussion of dualism and its consequences for development, see Hans W. Singer, "A New Approach to the Problems of the Dual Society in Developing Countries," *International Social Development Review*, 3 (1971), pp. 23-31.

<sup>33</sup> Aníbal Quijano, "The Urbanization of Society in Latin America," *Economic Bulletin for Latin America*, 13, 2 (1968), pp. 76-93. This article is not signed. However, it follows in general outline a paper by the same author, "Dependencia, Cambio Social, y Urbanización en Latinoamérica," *Cuadernos de Desarrollo Urbano-Regional*, 6 (March 1968), Santiago (CIDU, Universidad Católica de Chile).

<sup>34</sup> Jorge Hardoy, "Urban Land Policies and Land Use Control Measures in Cuba" (Report for the United Nations Centre for Housing, Building and Planning, 1970).

traditional (coöpted) elites and eager to assume a major governing role themselves. (In Peru, the military forces appear to have made these claims effective, though the results for development of the urban system remain unclear.)<sup>35</sup> It is noteworthy that the national 'counter-elite' of intellectuals is also the most receptive to modern technical and organizational innovation but sees its own aspirations for participation in governance thwarted by foreign powers and their national 'lackeys'.

An interesting case is that of Brazil, where the revolution which brought the military into absolute control of the country's governmental machinery may be interpreted, paradoxically, as the successful coöptation of the military – many of whose leading figures have strong provincial backgrounds by birth, education, and professional experience – by a national entrepreneurial establishment. Because unassimilated foreign elements constitute a relatively minor part of entrepreneurial groups in São Paulo and Rio de Janeiro, the military government has been able to pursue more nationalistic policies than would normally be expected under conditions of coöptation. These policies, however, have been directed more at problem areas that do not directly conflict with the central interests of the Brazilian business community, such as the building of trans-Brazilian highways and the colonization of new regions. Nationalistic efforts of this sort, as well as the absence of politics in the usual sense, have opened the door to the active collaboration of *técnicos* and intellectuals with the government and have all but destroyed potential counter-elites in the country. The long-term spatial effects of these new policies are likely to be spectacular.<sup>36</sup> They will contribute to the spatial integration of the Brazilian territory under conditions of internal dependency to the major core regions of the country.<sup>37</sup> But they will also uncover new possibilities for resources development, shift the gravitational field of the country's economic development away from coastal areas to the western frontier, and stimulate new urbanization along the major routes of interior penetration.

Under *accommodation*, a spheres of influence agreement of mutual non-interference may be tacitly reached according to which the manage-

<sup>35</sup> Eric J. Hobsbawm, "Peru: The Peculiar 'Revolution'", *The New York Review of Books*, December 16, 1971, pp. 29ff.

<sup>36</sup> See, for example, the extremely detailed study of new colonization along the Belém-Brasília Highway by Orlando Valverde and Catharina Vergolino Dias, *A Rodovia Belém-Brasília* (Rio de Janeiro: Fundação IBGE, 1967).

<sup>37</sup> For one of the most concise statements on the 'internal colonization' effects of the government's gigantic road building program, see Armando D. Mendes, "Um Projecto Para a Amazônia" (unpublished paper, Univ. Federal do Para, December 1971).



ment of the rural sector is left in the hands of the traditional governing elite, while the urban sector is 'turned over' to the innovative minorities to develop as they see fit, essentially as an enclave within the larger national territory. Enclaves of this sort are likely to be related more to the international economy (i.e., to the international urban system) than to the rural areas within the country. In some cases, such as Singapore, urban enclaves may be politically separated as well.<sup>38</sup>

This process of accommodation has been analyzed by Marcos Mamalakis in his theory of sectoral clashes.<sup>39</sup> Although Mamalakis' theory is expressed primarily in terms of major economic sectors (industrial vs. agricultural) it is easily translatable into spatial (regional) terms as well. In pre-Allende Chile, where the theory of sectoral clashes appears to be most strongly supported by the empirical evidence – though supporting data also come from Mexico and Argentina – the urban elites contained a large proportion of national minority groups of Germans, English, Yugoslavs, Jews, and Levantines (in addition to foreign, predominantly American, nationals), whereas the governing elite (the rurally-based 'oligarchy') was primarily of Spanish and Basque origin. Sectoral conflicts, reflected in the formation of political parties, had therefore certain ethnic-cultural overtones as well.

Finally, under conditions of *open hostility* events occur that lead to the disruption of existing relations of coöptation and/or accommodation. Conflict may assume a variety of forms, including *campaigns of national liberation* (Algeria), *the nationalization of foreign enterprise* (Cuba, Peru, Argentina, and Chile), *the elimination of ethnic minorities by either their physical destruction* (Jews in German-dominated Europe, Chinese in Indonesia) or *expulsion* (Indians from Kenya, French *colons* from Algeria), *economic pressure* (Chinese in Malaysia), *civil war* (Ibo in Nigeria), and *peaceful secession* (Singapore).

In some instances, the conclusion of hostilities has resulted in a renewed interest in rural development (involving the forceable transfer of resources from the core), with a consequent decline of growth in core areas and the concomitant renaissance of small to medium-sized provincial centers as base points for agricultural development.<sup>40</sup> Cuba provides perhaps the most clearcut evidence on this point, though a

<sup>38</sup> T. G. McGee, *op. cit.*

<sup>39</sup> Marcos J. Mamalakis, "The Theory of Sectoral Clashes," *Latin American Research Review*, IV, 3 (1969), pp. 9-46. In the same issue, see articles on Mexico by Barraza and Argentina by Merks.

<sup>40</sup> E. A. J. Johnson, *The Organization of Space in Developing Countries* (Cambridge, Mass.: Harvard University Press, 1970).

similar shift in allocation has also been reported for Malaysia.<sup>41</sup>

All of the situations discussed above relate to countries in which powerful innovative minorities in urban areas are *culturally distinct* from governing elites. But in other situations, such a split has not occurred and economic and political power is exercised conjointly from a dominant core region over ethnically and culturally varied populations (Northern Ireland, Soviet Union, Yugoslavia, Indonesia, Pakistan prior to the liberation of Bangladesh, Rhodesia, South Africa).

Where this occurs, the dependent regions will often claim to be 'oppressed' and generate political pressures for greater 'national' (i.e., regional) autonomy, ranging from complete secession to a number of 'protectionist' and 'federalist' solutions, including demands for 'preferential treatment.'<sup>42</sup>

Each of these solutions holds different implications for the development of urban systems. Some of them involve the massive transfer of populations (as has happened, most recently, in East Pakistan). Others lead to the isolation of the 'protected' areas from the 'virus' of urbanization (South Africa).<sup>43</sup> Still others produce vigorous urban-regional competition among federated states (Yugoslavia, India) with a consequent multiplication and strengthening of subsidiary core regions.<sup>44</sup> Occasionally, the mere threat of national independence or annexation to a neighboring country with similar ethnic traits may be sufficient to obtain preferential status (Commonwealth of Puerto Rico, French Quebec, South Tyrol).

These outcomes for urban systems may also be viewed from a perspective of (spatial) integration. The following table may help to recall the major patterns in this context. (The Roman numerals in the right-hand column refer to quadrants in Fig. 1; arrows indicate the principal direction of dominance.)

<sup>41</sup> James F. Guyot, "Creeping Urbanism in Malaysia," in Robert T. Dalaud (ed.), *Comparative Urban Research. The Administration and Politics of Cities* (Beverly Hills, Cal.: Sage Publications, 1969), ch. 4.

<sup>42</sup> Ivo D. Duchacek, *Comparative Federalism. The Territorial Dimension of Politics* (New York: Holt, Rinehart, and Winston, 1970).

<sup>43</sup> T. J. D. Fair, G. Murdoch, and H. M. Jones, *Development in Swaziland* (Johannesburg, Witwatersrand University Press, 1969). Also, L. P. Green and T. J. D. Fair, *Development in Africa. A Study in Regional Analysis with Special Reference to Southern Africa* (Johannesburg: Witwatersrand University Press, 1969).

<sup>44</sup> For Yugoslavia, see Ivo Babarović, *Regional Development Policies in Socialist Yugoslavia* (Unpublished Master in Regional Planning Thesis, Department of City and Regional Planning, Harvard University, 1966). For India, the concept of regional competition emerges from a study by Brian J. L. Berry, *Essays on*

| <i>Elite relationships</i>  | <i>Urban system</i>   | <i>Spatial integration</i>  |
|-----------------------------|---|---|
| 1. coöptation               | complete dominance of a passive periphery by the core: strong urban primacy; typical pattern of internal colonization, with strong linkages to international urban system   | integration based on dependency relationships and the continued imbalance of major urbanization processes (I → IV)  |
| 2. accommodation            | spheres of influence agreement leading to regional dualism: small number of modern urban enclaves relatively independent of traditional rural areas and joined more closely to international urban system than to national territory  | weak integration on basis of economic dependency: rural migrants belonging to national majority groups are prevented from reaching controlling positions in the urban economy occupied by innovative ethnic minority groups (I → III) |
| 3. open hostility           | if innovative urban minorities are effectively neutralized, the result may be a gradual transfer of resources from core to periphery, followed by accelerated rural development and the renaissance of small and medium-size provincial centers; development of a 'complete' urban hierarchy and attenuation of primacy | greater functional interdependency among regions and reduction of imbalances in urbanization: integration based on growing interdependency of urban centers (I → III)   |
| 4. regional 'protectorates' | policy of exclusion of urbanism from 'protected' areas or rural enclaves; core region dominance   | partial integration of urbanized (dominant) areas based on 'protected' labor pools in stagnant rural enclaves; economic dualism (I → IV)  |



- |                         |   |   |
|-------------------------|---|---|
| 5. federative solutions | preferential treatment and greater autonomy of 'associated states' and federal territories: competition among urban areas: emergence of subcores within each region | although frequently a fragile political arrangement, this 'solution' may eventually lead to a strong pattern of spatial integration based on urban-regional interdependency and the gradual attrition of peripheries: structured urban/regional competition (I $\longleftrightarrow$ II, I $\rightarrow$ III) |
|-------------------------|---|---|

*Dependency in Core-Periphery Relations:  
The Case of Chile*

In the first section of this paper, a basic distinction was drawn between the stock of potential power controlled by a person or an organization and the uses of this power in exchange relations with others. From the standpoint of empirical research, the latter is much easier to observe than the former. The process of exchange leaves visible traces and results in behavioral changes by at least one of the actors in the transaction. It is from a long series of such transactions that changes in the stock of power held by the participating actors may be inferred. In urban and, more generally, in spatial systems, the inferred distribution of power tends to be unequal, reflecting a dominant and persisting pattern of non-reciprocal exchange relations among cities and regions. I have called this the autonomy-dependency pattern and have argued that it will have a major influence on the relative growth and decline of cities whose economic and political fortunes are conjoined.

Code words such as city, region, or nation are useful for summing up exchange relations among a set of interdependent individual and organizational actors. Their use is permissible insofar as each refers to a relatively stable system of spatial relationships. Although integration may be achieved on a basis of either equality or dependency, the more interesting form, particularly in the case of newly developing nations, is the latter.

Spatial patterns of autonomy-dependency must be studied with respect to particular and limited domains of life. The processes of control by which dependency is secured are made effective through institutional arrangements which ensure a certain consistency of outcome. It is the spatial organization of these arrangements that allows us to extend the

concept of dependency into spatial analysis and to refer to control over particular domains of life in the periphery by organizational actors whose base of power is solidified within core regions.

In the following case study of dependency relations in Chile, I shall focus on those arrangements by which a core region centered upon the national capital assured its continued dominance over urban life in the rest of the country. Although the description is in the present tense, the reader should be aware that the facts cited pertain chiefly to the decade of the sixties. The new forces released by the Popular Unity Government since 1970 may bring about significant changes in the distribution of power and the spatial development pattern of Chile's economy.

Chile is an unusually good laboratory for the study of dependency relations. Its population is small and relatively homogeneous in ethnic origin, and cultural regionalism plays only a negligible role in national politics. Spatial integration, as measured by a shared historical past, a shared language and religion, a shared political system of great stability, and a well-articulated system of national transportation is exceptionally strong. At the same time, the overwhelming dominance of central power over even the minutiae of daily life is an acknowledged fact. For these reasons, Chile may stand as a classic instance of dependency relations in their purest form.

Five dimensions of the spatial organization of dependency will be described: municipal government, provincial administration, financial power, neighborhood power, and party organization. In the concluding section, some of the consequences of these patterns for the development of Chile's urban system will be considered.<sup>45</sup>

<sup>45</sup> For a descriptive account of the political system in Chile, see Federico G. Gil, *The Political System of Chile* (Boston: Houghton Mifflin Co., 1966), and James Petras, *Politics and Social Forces in Chilean Development* (Berkeley: University of California Press, 1969). Of particular value for an analysis of subnational development is Peter S. Cleaves, *Development Processes in Chilean Local Government* (Institute of International Studies, University of California, Berkeley, Politics of Modernization Series, No. 6, 1969), and John Friedmann (ed.), *Contribuciones a las Políticas Urbana, Regional, y Habitacional* (Santiago: Universidad Católica de Chile, Centro Interdisciplinario de Desarrollo Urbano-Regional, 1970), with contributions by Francis Earwaker, Rene Eyhéralde, Charles Frankenhoff, Ralph Gakenheimer, John Miller, Walter Stöhr, and Francisco Vázquez. For Marxist views of Chilean Development, see André Gunder Frank, *Capitalism and Underdevelopment in Latin America. Historical Studies of Chile and Brazil* (New York: Monthly Review Press, 1967), and Dale L. Johnson, "The National and Progressive Bourgeoisie in Chile," in James D. Cockcroft, André Gunder Frank, and Dale L. Johnson (eds.), *Dependence and Underdevelopment. Latin America's Political Economy* (New York: Anchor Books, 1972).

1. *Municipal Government.* The 277 municipalities of Chile are the only units of territorial government that stand between the individual citizen (or, more accurately, the extensive networks of familistic relations that form the texture of Chilean society) and the central authorities in Santiago. Elections for councilmen (*regidores*) are held every four years. The mayor is selected from the body of *regidores* by means of indirect elections, except that the mayors of Santiago, Valparaíso, and Viña del Mar are appointed directly by the President of the Republic.<sup>46</sup>

Municipal revenues, principally from the sale of licenses (*patentes*) for vehicles, dogs, moving picture theaters, mines, concessions, and horse racing, as well as business permits, are so low that most municipalities manage to do little more than pay their employees and monthly office bills. In 1967, the average municipal budget was only eight dollars *per capita*, but the amounts varied by size of municipality, from a low of three dollars for smaller units to nine dollars for municipalities with a population of over 100,000. Only the municipality of Santiago had a substantially larger budget, or nearly twenty-three dollars for each of its inhabitants.<sup>47</sup> A select number of municipalities receive additional income, as determined by national legislation, from local resource-using industries (wooden match manufacture, for instance) as well as from special taxes levied on ports, airports, and tourist facilities. But, in any event, the total amounts available for physical improvement and social welfare at the local level are insignificant. Local governments are responsible for the collection and incineration of garbage, for traffic control, street lighting, and public markets and gardens. As a practical matter, all other functions vital to the wellbeing of local inhabitants are directly managed out of national ministries in Santiago, including housing, public utilities, street paving, education, health care, social welfare, and public security.

This being so, municipal governments have little to do that is of any

<sup>46</sup> An ironic comment on the politics of local government in Chile comes from Peter S. Cleaves, *op. cit.*, pp. 13-14:

According to the Constitution, *alcaldes* or mayors are elected by the *regidores* from among themselves, except in the case of cities of over 100,000 inhabitants, where the president appoints an *alcade*. There is no constitutional stipulation that the presidential appointee must be a member of the elected municipal body. In 1969, there were thirteen cities in Chile with populations of over 100,000. However, since the Chilean Congress has avoided updating reapportionment since the 1930 census, *alcaldes* are appointed in only three cities: Santiago, Valparaíso, and Viña del Mar. Despite the low population of Viña in 1930, it was added to the list to facilitate tax supervision of its lucrative gambling casino.

<sup>47</sup> Cleaves, *op. cit.*, Table 4.

consequence. Having little to do, their employees are poorly paid. Being poorly paid, their professional quality is low. Restricted technical competence is then cited by central bureaucrats as the reason why local governments cannot be entrusted with greater responsibility.

It is generally agreed that the legal powers theoretically available to municipalities are not being fully exploited. One reason is that the national government consistently fails to transfer the full share of income taxes collected locally to which the municipalities are legally entitled.<sup>48</sup> Despite financial difficulties, a handful of municipalities has provided imaginative leadership in the provision of local services. But the institutional environment in Chile is inhospitable to displays of local ingenuity, and these exceptional experiences have not been imitated.

Local development is thus left almost entirely to the arbitrary judgment of Santiago officialdom. As a result, the fate of local populations is subject to all the vagaries of centralized state management, such as the limited attention span of key decision makers, their slow reaction time to new information, and the expediencies of national politics.

2. *Provincial Administration.* Following the French practice, provincial governors (*intendentes*) are appointed by the President of the Republic and report directly to the Minister of the Interior. Traditionally, the *intendente's* job has been to maintain 'law and order' in the provinces, provide political intelligence, and coordinate the work of the decentralized field offices of national ministries. Except for small emergency funds, *intendentes* have no development budgets of their own.

Since 1925, the Chilean constitution has included a provision for the election of provincial assemblies, but this has never been implemented. To regularize this anomalous situation, the formal powers of provincial assemblies to regulate municipal activities and control municipal expenditures were transferred to the *intendentes* in 1942. By this manoeuvre, the central government has been able to interject itself directly into issues of local governance. According to Peter S. Cleaves, this penetration has taken two forms:<sup>49</sup>

... independent servicing of the demands of the people and the coordination of interventions by the Minister of Housing and Urban Affairs and the Interior into functions that are directly under the jurisdiction of the municipality. To illustrate a recent trend in this direction: since 1965, the *intendencia* has had funds available for commun-

<sup>48</sup> *Ibid.*, p. 25.

<sup>49</sup> *Ibid.*, p. 31.

ity action while the municipality has continued to suffer from a lack of money. . . In recognition of the *intendencia's* capacity to take effective measures, *pobladores* (i.e., residents of poor quarters) and others have more and more bypassed the municipal structure to petition for direct government consideration of their problems.

In recent years, some efforts have been made to assign a greater developmental role to the *intendentes*. The National Planning Office (ODEPLAN) has divided the country's 25 provinces into ten development regions (plus a metropolitan district for Santiago) and has established small technical planning offices in the most important city of each region. These offices have done good work in recommending central budget allocations for their areas, but have not taken an active part in the implementation of specific projects or development programs. Exceptions to this are the provinces of Magallanes and Tarapacá, located at the extreme southern and northern ends of the country respectively, and far removed from the bureaucratic influence of Santiago. Because of Chile's interest in protecting these provinces against presumably covetous neighbors (Argentina, Bolivia, Peru), they have been given greater autonomy over their development than other regions. The city of Arica in Tarapacá, for instance, has been authorized to operate a municipal casino and to retain funds derived from its operation for local improvements. A technical staff, provided by the National Planning Office, has been assigned to work with the *Junta de Adelanto* of Arica (Arica's Development Junta) to steer the uses of these funds into growth-promoting investments. For Magallanes, a regional development corporation has been established with revenues from the extensive oil drilling operations in the province. The local branch of the National Planning Office is serving as a technical staff to the Corporation, and the local *intendente* has, in effect, become a regional development manager. Both these efforts appear to have been quite successful in stimulating local economic activity.<sup>50</sup>

Notwithstanding these regionally oriented planning activities, Chile's provinces remain politically and economically powerless, and their economic fortunes continue to be directed from Santiago. Although the information on which decisions concerning regional investments are made is better now than it used to be, the visible political pressures in Santiago (ministers can watch street demonstrations from their office

<sup>50</sup> Mariano Valle, *Planning Regional Development in Chile, Achievements and Perspectives* (MIT, SPURS, unpublished MS, 1969), and *The Planning Process in Chile* (MIT, SPURS, unpublished MS, 1970).

windows!) are generally more persuasive than the complaints of delegations from the largely 'invisible' provinces of Chile's periphery. During the 1960s, the major newspapers in Santiago typically buried provincial news on the inside pages. The periphery of the country was not considered especially newsworthy.

The general neglect of the provinces has left most provincial urban centers in the backwaters of the sprawling national metropolis.<sup>51</sup> Devoid of political power and without an economically prosperous hinterland, these cities have remained the passive objects of occasional national munificence. Local investments by the national government are therefore regarded as 'windfalls' and tend to generate only miniscule multipliers, since most of these are captured by Santiago. With 75 percent of all cities falling in the range of 5,000 to 40,000 inhabitants, none of Chile's cities, except for Santiago, have moved into self-sustaining growth. Unemployment in provincial centers has typically been two to three times the rate reported for the nation's capital, and it may be fairly assumed that most migrants arrive at provincial capitals, not because they expect to find a job there, but because living conditions in the nearby rural districts from which they come are even worse than in the city. Many migrants eventually move to Santiago.

3. *Financial Power.* In the late 1960s, Chile's public sector accounted for about three-fourths of all investments in the country. Development capital was channelled through a series of national corporations of which the most important was CORFO (Corporación de Fomento) which, in turn, controlled either wholly or in part a series of subsidiary enterprises. The headquarters of these and other national corporations (in housing, urban renewal, agrarian reform) were inevitably located in Santiago. Their capital was obtained partly from national revenues (including large-scale resource transfers from copper-producing regions) and partly from their own revenue-producing operations. CORFO, in addition, coordinated all major foreign loans.

<sup>51</sup> Public investments in Santiago increased from an average of 21.5 percent in the period 1960-64 to 31.5 in 1965-69. In the all-important housing and education sectors, the Metropolitan Zone of Santiago received 40.1 and 53.6 percent of all public investments in 1969, for a population that represented little more than one-third of the national total. As a result of these policies, Santiago had accumulated nearly one-half of Chile's regional product by 1970, over an area comprising only 2 percent of the national territory. See Sergio Boisier, *Polos de Desarrollo: Hipótesis y Políticas. Estudio de Bolivia, Chile, y Peru* (United Nations Institute for Social Development, Geneva, Report No. 72.1, January 1972), Tables 9, 17, and 18.



Although CORFO maintained several regional subsidiaries (the most important of which was CONORTE serving the northern provinces of Chile), these, too, were run from the center and had only a vestigial presence in their respective regions.

The remaining 25 percent of national investment was channelled through the private banking system. But even in the private sector (it might be more accurate to say, *especially* in the private sector), Santiago clearly dominated the scene. Provincial branch banks were tightly controlled by their parent banks in the capital and were permitted to make only very small loans on their own initiative. A number of so-called regional banks, such as the Banco de Talca, also operated out of Santiago. Carrying the name of their province, these banks were usually controlled by local landowners and provided a convenient channel for the transfer of funds to the center. The whole system worked to concentrate wealth, not only at the center, but also in the hands of the traditional 'oligarchy' whose principal families established their residence in Santiago.

This capacity of the private banking system to transfer capital resources from the periphery to the country's principal core region is a very common phenomenon. Unfortunately, it has received scant attention by students of regional development. Most of the funds collected through Chile's private banking system eventually found their way into Santiago real estate and into large commercial or industrial ventures, for the most part also located in Santiago. Private capital did not ordinarily seek out investment opportunities in the provinces. By tacit agreement, this job was left to CORFO.

4. *Neighborhood Power.* Chile's hypertrophied centralization was not unqualifiedly endorsed by everyone. Although most technicians relished it secretly, politicians of various political persuasions, but particularly Christian Democrats, mildly criticized the system for its cumbersome procedures, its lack of responsiveness to local needs, and its tendency to exclude from active participation in the national society the vast majority of the population. This phenomenon was labelled *marginalidad*, and it was partly in answer to this problem, that the reformist government of President Frei undertook to promote the regionalization of development policy.<sup>52</sup> The accelerated organization of *Juntas de Vecinos*

<sup>52</sup> The study of *la marginalidad* has occupied leading sociologists in Latin America for a number of years. The literature is surveyed by José Nun (Instituto Torcuato di Tella) in an article appearing in *Revista Latinoamericana de Sociología* (1969). The exact location of this reference is not available to me at present.

(neighborhood councils) was a second strategy pursued.

*Juntas de Vecinos* have a long history in Chile. As a rule, they spring into existence in the poorer urban barrios to mobilize the support of neighbors for petitioning the authorities to resolve some pressing local issue, such as paving a street, putting in a sewer system, or building a school. Once the problem is resolved, however, the Junta subsides into a dormant state until the next crisis. The Christian Democrats, after attaining to power in 1964, decided to use this traditional system for pointing the country more strongly in the direction of development.

Their original hope was to establish a national hierarchy of Juntas: the neighborhood Juntas of a given city would associate at the municipal level, federate at the provincial level, and confederate at the national level. The confederation of Juntas de Vecinos would then tie in with a National Ministry of Popular Promotion (*Promoción Popular*) that would represent the 'people's voice' (as filtered through the hierarchy of Juntas) at the cabinet level. Juntas were to be given full participation in local planning efforts and authorized to undertake certain public works on their own initiative.

Clearly, a system such as this could lend itself to a variety of purposes. A large number of proposals were floated, but only a few of them passed into legislation. The ideological commitment of the Christian Democrats was to mobilize the people for national development and so to transform popular discontent into a responsible voice. Or, as President Frei put it in his Second State of the Nation Message:<sup>53</sup>

My position in government has made me feel as never before *the necessity for the people to organize*; after the election, I was impressed by the fact that the centers of effective power, in many cases in the extreme minority, perpetuate a dead weight on the life of Chile which is unacceptable. For this reason, the silence of a disorganized people must transform itself into a *responsible voice*. And therefore, *Promoción Popular* is not just an idea or even an electoral platform for us; it is the profound necessity to *transform in an organic manner the basic structures of our society* [emphasis included].

But the political opposition saw matters in a different light. They suggested a devious manoeuvre by the party in power to consolidate its political bases in the local barrios. A National Council of Popular Promotion survived the determined legislative opposition and, temporarily attached to the President's Office, continued its organizing work among the Juntas and other popular organizations, such as Mothers'

<sup>53</sup> Cited from Cleaves, *op. cit.*, p. 43.



Centers and Sport Clubs. But the Council no longer held a monopoly over community action. The parties of the left (Communists and Socialists) pursued the work of organizing local Juntas with equal vigor. At the same time, high-level officials in several government ministries (especially Housing and Agriculture) promoted their own versions of 'maximum feasible participation'.

This attempt to mobilize neighborhood power and harness it to the goals of national development had no immediately visible effects on Chile's urban system. At one point, it had been the intention of the Christian Democrats to create in the *Juntas de Vecinos* a system of power paralleling that of municipal governments, evidently in hopes of ultimately replacing the latter. (Municipal Councils tended to be controlled by members of parties in opposition to the Christian Democrats.) This hope, however, was effectively torpedoed. Nor was the national distribution of power significantly affected by this effort to mobilize the population of local communities and, particularly, the poor. Control over even local investments remained centralized and relations in ownership and production were left untouched.

5. *Party Organization.* Political parties in Chile reflect sectoral and class interests more than interests that are purely regional or local. During the 19th century, a major split divided rural-based Conservatives from urban-based Liberals, with the former dominated by large land-holding families and the latter representing the new commercial-industrial elites of Santiago. With the further growth of the national core area, however, additional urban parties came into existence, each finding support among incipient social groups. White collar office workers and small shopkeepers became the mainstay of the Radical Party (Chile's Grand Old Party), Communists entrenched themselves among organized workers (initially in the copper mining camps of the North, but eventually shifting the base of their operations to the major cities in the center). Socialists (a relatively elitist group in Chile) found special favor among intellectuals and managed to win additional support in Chile's frontier areas, principally Magallanes, while the Christian Democrats appealed primarily to the new class of *técnicos*, recently enfranchised women voters, and broad sections of the relatively small, well-to-do middle class.

In national elections, parties would typically nominate their most trusted militants to run for political office from one of the provincial districts. For a brief, intensive period thereafter, the candidate would present

himself to his local electorate. Once voted into office, however, the new *diputado* or *senador* from Province X would rush back to Santiago (where, in any case, he resided) to throw himself into the battle of national politics which was aligned principally by loyalties to a political party and strong, personalistic leaders. They were in no way expected to 'represent' local interests in the American sense.

Despite outward appearances of contest, mutual accommodation succeeded for a long time in governing relations between rural and urban parties, in the sense that neither would seriously interfere in the interest spheres of the other. Agricultural policy was thus left to the Conservatives, while urban policy became the battle ground of a number of mutually antagonistic urban parties.

With progressive urbanization, however, the Conservative Party lost ground and eventually transformed itself into yet another urban party that would form frequent alliances with other political parties in a conservative coalition.

In their ascendancy to power, the Christian Democrats had to compete against an opposition of Radicals, Communists, and Socialists whose main strength was in Santiago and other large cities (Antofagasta, Concepción). This left the Christian Democrats chiefly with the rural periphery of the country, traditionally dominated by the Conservatives. As a result, the Christian Democrats became the first Chilean party actively to solicit the support of regional populations. It was this regional orientation which subsequently led the Party to advocate a more vigorous agrarian reform program than the preceding government of Jorge Alessandri had done and to promote the regionalization of development programs generally. But the effort was only half-hearted. It failed to bring about a drastic change in the distribution of governmental and economic power, and the earlier tendencies toward continued concentration of economic activities and population in Santiago continued.

The foregoing description of the spatial element in national politics is, of course, a highly simplified one. Not only did Chile have minor regionalist parties from time to time, such as PADENA, but the North of the country, with its heavy concentration of employment in copper-mining, often played a decisive role in determining the outcome of national elections. Similarly, regional concentrations of national minorities (Germans, Yugoslavs) have occasionally influenced voting behavior in certain parts of the country. Nevertheless, spatial considerations were definitely subordinate to ideological and personalistic issues. One did

not customarily rise to national prominence from local office; indeed, as I have tried to suggest, neither municipalities nor provinces provided much scope for political talent. Anyone who wished to succeed politically, had to move to Santiago.

The absence of political regionalism in Chile helps account for the sense of strong national unity. National politicians in Santiago generally thought in national terms (as well as sectorially and along class lines). The system suffered from an excessive factionalism which at times was carried to an extreme of personal bitterness, but the formation of regional blocks, so common in other countries, was not part of the picture. An appearance of regionalism is given by the geographic distribution of major economic sectors: miners in the North, industrial workers and middle class in the cities of the center, large-scale agriculture in the central Valley of Chile, and *minifundos* (small-scale farming) in the South. Yet, to this day, it is the *national* point of view that prevails. And, invariably, the national point of view is that of Santiago.

6. *Conclusion.* It was the intent of this section to describe some of the ways in which relations of dependency are organized in Chile's urban system. In the absence of a strong countervailing power having a regional territorial base, most of Chile's cities in the shadow of the Santiago metropolis have been reduced to penury. An incipient active periphery was observed at the two extremes of the national territory, where the relatively exposed border positions of both Magallanes and Tarapacá have led to modest increases in their autonomy to make development decisions. But elsewhere, local populations wait more or less passively for the decisions of superior authorities, though they will occasionally seek to influence these decisions through appropriate political means. Initiative is at low ebb, and those who wish to find more scope for their gifts of enterprise, eventually relocate in Santiago.

The resulting ascendancy of the national capital in Chile's urban system is best illustrated by the following table, showing the historical evolution of three indices of primacy (the ratio of Santiago's population to the next three lower-ranking cities in combinations of two, three, and four). Regardless of which index is chosen, the relative weight of Santiago's population has been continuously increasing and, during recent decades, at an accelerating rate.<sup>54</sup>

<sup>54</sup> John Friedmann, *Urbanization, Planning, and National Development*, *Op. cit.*, chapter 5, Table 3.

Table 1. *Chile: Index of Primacy, 1865-1960*

|              | 1865 | 1875 | 1885 | 1895 | 1907 | 1920 | 1930 | 1940 | 1950 | 1960 |
|--------------|------|------|------|------|------|------|------|------|------|------|
| 2-city index | 1.64 | 1.31 | 1.72 | 1.92 | 1.76 | 2.33 | 2.94 | 3.45 | 4.44 | 5.18 |
| 3-city index | —    | 1.08 | 1.36 | 1.40 | 1.28 | 1.67 | 2.11 | 2.40 | 2.82 | 3.18 |
| 4-city index | —    | —    | 1.16 | 1.18 | 1.11 | 1.42 | 1.82 | 2.13 | 2.50 | 2.77 |

*Final Comments*

The influences on the development of urban systems are multiple and reciprocal. Governmental and economic power make up only one set of such influences, though I will argue that it is probably the most important set. In practice, of course, these influences are difficult to isolate, and it is still more difficult to measure their effects within a cybernetic framework of analysis. The customary two-variable regression models, dear to economists and sociologists, fail to yield the insights one would want. This suggests that we may have to choose between quantitative precision in research, leading to very restricted and possibly misleading insights, and a more comprehensive, qualitative approach which may have to sacrifice the elegance of mathematical formulations for a deeper historical and conceptual understanding. If we should opt for the second approach, formal models of the sort introduced at the start of this paper serve primarily a heuristic purpose. They fall short of theoretical constructs that purport to model complex causal relations in the development of urban systems. On the other hand, they focus the attention of researchers on critical variables and their interrelationships and help in posing questions that may eventually lead to significant insights into the workings of historical processes. They are preliminary to empirical research.

I have placed the emphasis in this paper on the spatial distribution of governmental and economic power. It could be argued, as William Alonso has done, that the degree to which all national policies have implications for urban growth, sectoral rather than spatial distributions of power are the critical variables.<sup>55</sup> Most probably, both viewpoints are correct and complement each other. Whatever the conclusion, however, the model I have proposed points to a series of fascinating questions about the effect of power relations on the structural growth of

<sup>55</sup> William Alonso, "Problems, Purposes, and Implicit Policies for a National Strategy of Urbanization" (Working Paper no. 158, Institute of Urban and Regional Development, University of California, Berkeley, August 1971).

hierarchical systems. It is with a view to clarifying these questions and setting the stage for a large-scale research effort that the present paper has been written.

**Growth in Subnational Regions**

\_\_\_\_\_

\_\_\_\_\_

# 12

## Natural Resource Endowment and Regional Economic Growth

Harvey Perloff and Lowdon Wingo, Jr.

WE ARE CONCERNED in this paper with the relationship between the natural resources within the various parts, or regions, of a country and what might be called the geography of national economic expansion.

The development of the economy of the United States provides a case study of resources in space interacting with the other elements of economic growth that is especially illuminating: (1) It covers an extended spectrum of growth from early agricultural beginnings to status as an advanced, industrial-and-service-oriented economy, thus affording an opportunity to examine the role of resources in different stages of national economic growth. (2) It covers a wide variety of regional resources and growth situations, providing a rich set of variations on the interactions of the national economy with its geographic components. And (3) it can be examined with the help of a wealth of historical information and statistical data. Specifically, Resources for the Future has concluded a three-year study of the regional characteristics of the growth of the U. S. economy. The results of this effort have been made available in a book entitled *Regions, Resources, and Economic Growth*.<sup>1</sup>

<sup>1</sup> Harvey S. Perloff, Edgar S. Dunn, Jr., Eric E. Lampard, and Richard F. Muth: *Regions, Resources, and Economic Growth* (Baltimore: The Johns Hopkins Press, 1960). We owe a great deal to Messrs. Dunn, Lampard, and Muth; we wish to absolve them, however, of responsibility for any errors or inadequacies in this paper.

The discussion of growth here will be limited to changes in what we refer to as the *volume* of economic activities (e.g., increases in population, employment, value added, and the like), acknowledging that this is but one facet of economic growth. Growth as defined by changes in *welfare* (changes in per capita income, for example) is discussed in Part V of *Regions, Resources, and Economic Growth*; changes in state per capita are analyzed in some detail by Simon Kuznets in "Industrial Distribution of Income and Labor Force by States, 1919-21 to

Reprinted from *Natural Resources and Economic Growth*, ed. Joseph J. Spengler (Washington, D.C.: Resources for the Future, 1961), pp. 191-212, with permission of the authors and Resources for the Future, Inc.



This paper draws heavily on the research that went into this book and on the conclusions that emerge from it. Given this storehouse of materials, we have been tempted into speculation on the broad relationships involved in the resources-growth problem.

#### RESOURCES AND GROWTH IN A BROAD HISTORICAL FRAMEWORK

One of the insights emerging from an examination of the history of American economic development is the difficulty of defining "resource endowment" in any long-run, substantive sense. In the short run, endowment is simply the inventory of those natural materials that are required in some degree by the national economy responding to internal consumption demands and to its position in international trade. As the requirements of the economy change, the composition of the inventory shifts, and in this sense "resource endowment" is a changing concept closely associated with the dynamics of economic growth. In short, the answer to what constitutes "resource endowment" is rooted in the determinants of final demand—consumer preferences and income distribution, as well as foreign trade—on the one hand, and in the current organization and technology of production on the other. As these variables change, so will the content of resource endowment. And, clearly then, as the composition of resource endowment changes, there will tend to be substantial changes in the relative advantage among regions supplying material inputs (and services) for the national economy.

The impact of these shifts can be sketched in with broad strokes by identifying stage by stage what have been the "natural resources that count" in the national economy. This requires us to tell again a familiar story, but with a special focus.

#### *The Early Agricultural Period*

From its colonial origins the American economy developed as a producer of resource inputs into the rapidly expanding European economy. To serve such a function the endowment which counted in early America was arable land with its environmental complements of climate and water, and this, with access to the growing European market for

---

1955," Part III of "Quantitative Aspects of the Economic Growth of Nations," *Economic Development and Cultural Change*, Vol. VI, July 1958.

We assume that the broad features of regional economic expansion in the United States are familiar to our readers, and so we have presented no figures to describe them here. Those interested in such detail are referred to Chapters 2 and 3 of *Regions, Resources, and Economic Growth*.

agricultural staples, set up the conditions for regional growth in early America. It was quite logical, hence, that the regional economies developed a certain archetype: a good deepwater port as the nucleus of an agricultural hinterland well adapted for the production of a staple commodity in demand on the world market.

The growth potential of these nucleated regions depended heavily on the extent and "richness" of the hinterland accessible to port. Since good agricultural land was almost a free resource while labor and capital were dear, the expansion of production was effected by bringing more land into production and so extending the limits of the hinterland. Much of early American history is dominated by the great rivalries for control of hinterland that emerged between New York and Boston, Philadelphia and Baltimore, Charleston and Savannah. This expansion of the hinterlands took place through social overhead investment in transportation facilities beginning with the Massachusetts road system in the seventeenth century,<sup>2</sup> later producing the Erie Canal, and finally motivating the half century of railroad construction stretching from the Baltimore and Ohio's first crude line reaching out to the rich wheatlands in Maryland and Pennsylvania to the driving of the golden spike at Promontory Point, Utah.<sup>3</sup> The force of the outward push for land is suggested by the fact that population west of the Alleghenies, which was estimated by the 1790 Census at 109,000, by 1840 had become almost six and a half million, with more than 87 per cent of the labor force involved in agriculture.

Even though the data on this period are not very satisfactory, we can draw these general conclusions: (1) the *regional* endowment that made for growth was "good" land advantageously situated with respect to the market centers; (2) the distribution of economic activity in the period before 1840 was essentially a function of the expanding, nucleated, agricultural regions reaching into the economic vacuum of an unsettled continent to bring ever greater areas of land under cultivation; and (3) this resource-dominated expansion of the economy set the stage for the next important development by establishing a geography of markets, transport, and labor force to condition the nature of succeeding growth.

<sup>2</sup> The importance of hinterland in the growth of the early centers is vividly described by Carl Bridenbaugh, *Cities in the Wilderness* (New York: Ronald Press, 1938).

<sup>3</sup> A detailed treatment can be found in Paul H. Cootner, "Transportation Innovation and Economic Development: The Case of the U. S. Steam Railroads." Unpublished Ph.D. thesis, MIT, 1953.

*The Minerals-Dominant Economy*

Somewhere around 1840-1850 the next important resource stage began—as a result of the emerging minerals-dominant economy. The rapid growth of the railroads and the expansion of processing industries resulted in new input requirements: a new set of resources became important and a new set of locational forces came into play. The first part of this period was dominated by the growing demand for iron and steel and by the rapid elaboration of their production technology.<sup>4</sup> At this point it was the geographical juxtaposition of coal, iron ore, and the market which afforded the great impetus for growth. The importance of minerals, unlike agricultural land, was not alone in their direct contribution to regional growth so much as it was in the nature of their linkages with succeeding stages of production. It was not so much the mining of coal and iron that was important for growth, as the making of iron and steel products, which could not be separated from the sources of its mineral inputs. The early concentration of steel making in western Pennsylvania was a result of these relationships, for this area was not only well endowed with deposits of iron ore and coal but was central to a concentrated market stretching from Boston and New York westward. As the center of gravity of the market shifted west and as Mesabi ores replaced depleted local ores, the iron and steel industry also shifted westward along the southern shores of the Great Lakes.

With the increase in the demand for nonferrous metals, the depletion of accessible ore deposits in the East and the penetration of the West by the railroad net, a new role in regional growth was played by mineral resources endowment. In the Mountain region stretching from the Canadian border to the Southwest states, the mining of metal ores was the lead factor in economic development: in 1870 when mineral extractions involved the employment of 1½ per cent of the labor force nationally, in the Mountain states it accounted for no less than 26.54 per cent, after which it declined until in 1950 the proportion was 3.44 per cent, still twice as much as the national average. Except for primary processing of ores, however, this resource base did not induce the location of any substantial amount of linked activity in the Mountain states. With most of the weight loss taking place during concentration and smelting, the distribution of the market governed the location of succeeding stages of metals fabrication, and the major markets were concentrated in the Northeast.

<sup>4</sup> In 1880, some 70 per cent of steel output went into rails. Cootner, *op. cit.*, Chapter V, pp. 13-14.

The extent to which changes in both demand and supply conditions influenced regional resources activities is suggested by the data in Table 1, showing figures for interregional production shifts for pig iron, copper and lead. Several points are worth noting: (1) the period of great growth in the output of these mineral products (1870-1910) corresponds with the most extensive interregional shifts in their production; (2) truly huge shifts in lead and copper production took place from the Great Lakes region to the Mountain states (and to a lesser extent to the Southwest and Far West) during this period; and (3) there has been a steady shift in pig iron production throughout the entire period 1870-1950 from the Middle Atlantic states to the Great Lakes and, to a lesser extent, the Southeast. These data underline the highly selective regional effects resulting from the growth of a mineral-based economy.

Some notion of the extent of changes in national requirements of material resources is provided by a measure of the changes in the composition of the value of purchases in constant dollars within the broad mineral categories over the period 1870-1950. (This measure is the same as that employed in Table 1; namely, the end of period total which would have to be redistributed among classes to recreate the beginning of period percentage distribution.) Thus, within the mineral fuels the total shift over the period was equivalent to 57 per cent away from bituminous and anthracite coals and toward petroleum and natural gas fuels, that is, towards materials which would hardly have been considered as resources ten years before this period. Among the metals during the same period the total shift was almost 34 per cent, away from iron, lead, and tin and in the direction of the light metals and ferroalloys—one-fourth of this shift has been in the direction of metals for which the economy of 1870 had little or no use, such as aluminum, manganese, nickel, and molybdenum. Finally, among the nonmetals (and here the availability of data limits us to the period 1910 to 1950) the internal composition in this shorter period shifted by 31 per cent, away from stone and toward other construction materials, as well as toward basic chemical materials. Something more than one-fourth of the shift was to materials which would not have been considered as resources in 1870.<sup>5</sup>

An especially important instance of the regional effects of changes in national requirements is provided by the case of petroleum and natural gas in the Southwest in recent decades. Here the effect resulted

<sup>5</sup> Calculated from data in a forthcoming RFF study by Neal Potter and Francis T. Christy, Jr., "U. S. Natural Resource Statistics, 1870 to 1955" (Preliminary Draft, with revisions to 11/1/59), Tables McT-22, 23, 24, 33, and 35.

TABLE 1.—PERCENTAGE CHANGE IN REGIONAL DISTRIBUTION OF U. S. PRODUCTION OF PIG IRON, COPPER AND LEAD, 1870-1910 AND 1910-1950

| Region <sup>b</sup>              | Pig iron                      |           | Copper                        |                        | Lead                          |           |
|----------------------------------|-------------------------------|-----------|-------------------------------|------------------------|-------------------------------|-----------|
|                                  | Percentage Shift <sup>a</sup> |           | Percentage Shift <sup>a</sup> |                        | Percentage Shift <sup>a</sup> |           |
|                                  | 1870-1910                     | 1910-1950 | 1870-1910 <sup>c</sup>        | 1910-1950 <sup>c</sup> | 1870-1910 <sup>c</sup>        | 1910-1950 |
| New England..                    | - 1.62                        | - .06     | - 6.90                        | 0                      | - .41                         | 0         |
| Middle Atlantic                  | -15.98                        | -15.65    | - 1.44                        | - .08                  | - 1.02                        | + .34     |
| Great Lakes...                   | +11.13                        | + 7.37    | -62.36                        | -17.72                 | -58.89                        | - .45     |
| Southeast.....                   | + 4.90                        | + 5.80    | - 6.39                        | - 1.57                 | - 3.89                        | + .75     |
| Plains.....                      | 0                             | 0         | 0                             | + .33                  | + 7.09                        | - 7.89    |
| Southwest.....                   | 0                             | 0         | +27.73                        | +23.77                 | + 1.63                        | +10.30    |
| Mountain.....                    | 0                             | 0         | +39.49                        | - 2.28                 | +54.61                        | -10.17    |
| Far West.....                    | + 1.57                        | + 2.54    | + 9.86                        | - 4.03                 | + .88                         | + 7.12    |
| Total Shift....                  | 17.60                         | 15.71     | 77.08                         | 24.10                  | 64.21                         | 18.51     |
| Percentage growth of output..... | 1,538%                        | 115%      | 2,727%                        | 67.8%                  | 1,983%                        | 11.7%     |

<sup>a</sup> The end-of-period percentage of total national production less the beginning-of-period percentage. Thus, with respect to pig iron in the period 1870-1910, New England percentage share of national production was 1.62% less in 1910 than it was in 1870. The figure for Total Shift (sum of the absolute value of the shifts  $\times \frac{1}{2}$ ), then, represents the percentage of total national production which would have to be redistributed in order to recreate the beginning-of-period percentage distribution by regions.

<sup>b</sup> The states composing the regions are as follows: *New England*—Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut. *Middle Atlantic*—New York, New Jersey, Pennsylvania, Delaware, Maryland, District of Columbia. *Great Lakes*—Ohio, Indiana, Illinois, Michigan, Wisconsin. *Southeast*—Virginia, West Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana. *Plains*—Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas. *Southwest*—Oklahoma, Texas, Arizona, New Mexico. *Mountain*—Montana, Idaho, Wyoming, Utah, Colorado. *Far West*—Washington, Oregon, California, Nevada.

<sup>c</sup> The 1870-1910 shift figures for lead and copper should be viewed with caution. The 1870 figures upon which the shifts are based represent the regional composition of the *current dollar value* of domestic mine production (the only data which were available), while all later figures used express domestic mine production in *short tons*. Since we are dealing with regional composition, the resulting shift figures would be seriously compromised if there were substantial price differentials at the mine head among the major producing regions in 1870. For the purposes of the discussion following, it is assumed that such price differentials would exert at worst a modest influence on the 1870-1910 shift figures.

<sup>d</sup> Excludes unallocated production in Pennsylvania, Tennessee, and Vermont of 1.59%.

(For sources see p. 7)

not only from a powerful, direct mining leverage (as was noted in the case of the Mountain region), but also from the availability of a cheap, convenient fuel which altered substantially the region's relative advantages for certain classes of industry. The happy coincidence of these mineral fuels with rich deposits of salt and sulfur provided a resource base for a rapidly expanding chemical industry. Thus, petroleum and gas extraction and refining, responding to a huge and growing national demand, served to change the economic conditions of production throughout the entire Southwest.

Summarizing the broad sweeps of the period of very rapid growth from about the middle of the nineteenth century, we note that during the first half of this period (to the end of the nineteenth century) there were two great overlapping resource effects conditioning the subnational distribution of economic activity: (1) geographically the more widespread effect was that of agriculture continuing to spread out over the arable lands—as in the early period of economic development, but pulling with it an increasing component of processing and servicing activities; and (2) the developmentally dominant effects emerged from the growth of the minerals economy, shifting rapidly among regions, triggering, intensifying, or transforming the nature of regional growth patterns.

The second part of this modern period—that is, the first half of the twentieth century—has been largely characterized by an elaboration and “deepening” of the subnational economy building upon the geographic pattern of activities brought about by the great interregional resource shifts of the nineteenth century. Resource activities declined in relative importance in the national economy throughout the period, but their real importance lay in the role they had played historically in defining the economic basis for the succeeding stages of regional growth—in the movements of population and industry among the regions. In a very real sense the classical resource effects were playing themselves out, as the service sector moved into a dominant position and as technological and other changes (such as price changes which made recapture of waste products economical) brought about a long-range

---

SOURCES for Table 1: Shift figures computed from Perloff, Dunn, Lampard, and Muth, *Regions, Resources, and Economic Growth* (Baltimore: The Johns Hopkins Press, 1960), T. 75, p. 205; T. 76, p. 208; T. 77, p. 210. Growth of output computed from the long-term series in a forthcoming RFF study by Neal Potter and Francis T. Christy, Jr., “U. S. Natural Resource Statistics, 1870 to 1955” (Preliminary Draft, with revisions to 11/1/59), Lead, Table MT-14; Copper, Table MT-12; Pig Iron, Table MT-28.

reduction in the proportion of raw materials to total output,<sup>6</sup> thus weakening the linkages of economic activities to their resource inputs. The power of the "market magnet" loomed as the dominant locational force operating in the economy.

*The "Services" Era and "Amenity Resources"*

By mid-century, moreover, an additional resource effect was beginning to influence the distribution of economic activity among the regions. To understand the importance of this effect requires us to move away from a definition of resource endowment which sees resources exclusively as tangible materials upon which technology works in the production of goods, and toward one which sees natural resources as including other features of the natural environment which have consequences for economic decisions. Natural resources, then, need not *enter* directly into the processes of production, but only *influence* directly the location of markets as well as of production. This broader definition embraces a group of physical environmental conditions which we will refer to as the "amenity resources"—that special juxtaposition of climate, land, coastline, and water offering conditions of living which exert a strong pull on migrants from less happily situated parts of the nation.

This amenity-resource effect derives from the interplay of a number of developments within the national economy and society. First, there is the growing importance of the nonjob-oriented, as well as the job-seeking, migrant. Some 8 per cent of the U.S. population is over 65 years of age, and the proportion of this age group in the total is growing. Approximately two-thirds of these persons are not working and many enjoy some form of paid retirement. Since most consumption items can be acquired with only minor interregional differences, many of these persons will seek out the more intangible resource services, such as climate and coast, that do have substantial interregional variations.

Another important development is the growth in the number and

<sup>6</sup> The Potter-Christy data indicate that between 1870 and 1955, when real GNP expanded 16 times, the output of the resources industries expanded only 5½ times. In terms of output (in 1954 prices), the extractive industries dropped over this period from ⅓ of GNP to 12 per cent. The greatest declines were in the products of forestry, fishing, and agriculture. Output in mining rose as a percentage of GNP until the 1920's and since has shown a moderate decline relative to GNP. Kindleberger finds a similar trend in the declining relative use of raw materials in Europe. Charles P. Kindleberger, *The Terms of Trade: A European Case Study* (Cambridge and New York: Technology Press and Wiley, 1956), Chap. 8, pp. 176-212.



significance of industries whose ties to resource inputs and national market centers are relatively weak. These are the so-called "foot-loose" industries which are distinguished from other industries in the fact that they have an unusually broad spectrum of locational alternatives available. Such an industry may be labor-oriented in terms of requiring unskilled or semi-skilled labor, such as the apparel industries, or in terms of a highly technical labor requirement, such as the research and development industries. It may be climate-oriented, as in the case of the aircraft industries. Or it may be an industry whose unit transportation costs are negligible in terms of the value of the product, such as instrument and optical goods producers. All of these have in common an array of locational possibilities which permits them to settle in amenity-rich areas without doing violence to the economics of their activities. The growth of the transportation equipment industry (mainly aircraft) in California is an excellent example. During the period 1939-54 California realized some 35 per cent of the national shift in employment in the industry, and this accounted for a very large share of California's total increase in manufacturing employment.

Finally, there is the effect of a rising per capita income throughout the nation. Given the high elasticity of demand for travel and recreation, rising incomes have meant an increasing export market for regional amenity resources in the form of tourist services to vacationers.

Even before mid-century the great shift in population was in the direction of states that had advantages in these amenity resources: Florida, the Southwest and the Pacific Coast states. During the 1940-50 Census period, this great arc of states stretching from Florida on the southeastern rim to Washington on the northwestern rim<sup>7</sup> (which contained 16 per cent of the national population in 1940) absorbed some 40 per cent of the total increment of national population growth.<sup>8</sup> The movement in the direction of the amenity resources is strong, and even though we are not certain how much of the movement to specific regions can be attributed directly to this resource influence and how much to

<sup>7</sup> Florida, Texas, New Mexico, Arizona, California, Oregon, Washington.

<sup>8</sup> In the case of Florida, between 1940 and 1950 the native white population of the state increased by 54 per cent, adding an increment of 707,300 to the 1940 population of 1,304,000. Of this increment the increase in Florida-born residents accounted for 210,000, while the increase of residents born in other states accounted for 497,300, some 251,600 of which were born in states north of the Ohio River. New York alone accounted for 30 per cent of these. Everett S. Lee, Ann R. Miller, Carol P. Brainard, and Richard A. Easterlin, *Population Redistribution and Economic Growth, United States, 1870-1950* (Philadelphia: The American Philosophical Society, 1957), Table P-3, p. 257.



other factors, given a highly mobile population with rising incomes and retirement payments, it seems fairly certain that the direct influence of the amenity resource will increase rather than diminish.

And so, in the broad perspective of history, the changing content of resource endowment has had a succession of effects in the inter-regional distribution of economic activity. As "new" resources moved to the forefront of the national economy, new advantages for economic growth were created for those regions well endowed. This much seems certain: in terms of the distribution of national economic activity over the landscape, resource endowment has mattered a great deal.

### RESOURCES AND THE MECHANICS OF REGIONAL GROWTH

Regional growth typically has been promoted by the ability of a region to produce goods or services demanded by the national economy and to export them at a competitive advantage with respect to other regions. We have already referred to three such cases touching upon resources—the leverage of minerals in the growth of the Mountain states, of petroleum and natural gas in the growth of the Southwest, and of amenity resources in the growth of Florida. The role of timber in the development of the Pacific Northwest and the role of agricultural commodities in the development of the Plains states are equally instructive. This ability to export induces a flow of income into the region which, through the familiar multiplier effect, tends to expand the internal markets of the region for both national and region-serving goods and services. The extent of the multiplier effect is related to certain "internal" features that characterize the economic and social structure of the region. Regions tend to differ substantially in the degree of development that becomes associated with the growth of the export industries and in what happens to the income that flows in from the export sales.

Some of these internal features are related to the nature of the export industries and particularly to the localized industrial linkages, and services attaching to the export sector are also important here. Thus, for example, it has been noted by historians that the shipment of *heavy* export products from a region has influenced the development of substantial transportation facilities and services within the region. The quantity and type of labor required by the export industries and relative levels of wages paid has, of course, an obvious relationship to the "internal" development of a region. Another important feature is the income distribution that tends to be associated with a given type of regional export product. Douglass North has pointed to the differential

effect on regional development in the nineteenth century of the plantation system in the Southeast for the production of cotton and tobacco—with its highly unequal distribution of income, as compared to the independent-farmer production system of the Midwest—with its broad income base and its growing markets for local goods and services.

“Internal” regional development takes the form both of internal structural changes (such as an increase in the proportion of the labor force employed in manufacturing and service industries) and an expansion of the local market for all sorts of goods and services. As the regional market expands and region-serving activities proliferate conditions may develop for self-reinforcing and self-sustaining regional growth, and new internal factors may become important in determining the rates of regional growth, such as external economies associated with social overhead capital and the agglomeration of industries, and internal economies of scale. At any rate, the occurrence of rapid self-sustaining growth involves a shift in the relative importance of growth factors—away from the dominance of the export sector and in the direction of the internal organization of production—which makes it possible for the region to play a more elaborate role in the national economy. This highly simplified exposition of the regional growth process needs to be hedged with many reservations, but it brings to the fore the context within which the effects of resource endowment play out their role.

The “export” and “internal” determinants of regional economic expansion can be brought together in the concept of *cumulative advantage*. But any advantage which a region may have *vis-à-vis* other regions is, of course, always relative. This is so whether the focus is in terms of input and market advantages in the production of a single product or the products of a single industry, or whether the focus is in terms of cumulative advantages for over-all economic growth.

The conditions making for relative advantage can be of many sorts. Given our focus on the role of natural resource in growth, it is suggestive to view relative advantage as resource-based and nonresource-based. As already noted, resource-based advantages have afforded the conventional route to regional economic growth in the United States. In terms of their consequences for regional economic expansion, resources can be described as “good” or “bad” depending on their capacity to provide a vigorous economic linkage with the national economy and to extend the internal markets of the region. A good resource for a region can be identified, first, by its ability to support an extensive stream of nationally-wanted production. Here attention is focused on the characteristics of the national demand curve for the resource and the relation-

ship of the region's supply conditions to those of the other regions: these must afford a substantial promise that production of the resource in the region will expand. In short, the demand for the resource must be derived from final and intermediate demand sectors of the national economy exhibiting a high income-elasticity of demand. Secondly, production of the resource must be characterized by extensive locationally-associated forward and backward linkages. And, finally, the resource must be characterized by a high regional multiplier—that is, a substantial proportion of the returns from the export sector must find its way into active demand for regionally produced goods and services.

Thus, a region's resource endowment is "good" to the extent that it is composed of resource products which rate high by these criteria. A "poor" resource endowment is one whose potential for inducing growth is, accordingly, not very high. In general, the importance of resource endowment in regional growth derives from its ability to alter the region's over-all cumulative advantage position. This will vary among regions and among resource components, and especially will it vary over time as shifts in the composition of rapid and slow growth sectors of the national economy change the bill of inputs.

Most agricultural products rate low on this "growth" scale. The agricultural sector in recent decades has expanded at about the rate of population growth<sup>10</sup> and its products have had an income-elasticity of much less than one.<sup>10</sup>

Thus, taken in aggregate, agriculture will rarely make much of a contribution to the cumulative growth advantages of a region, except in the case of a region whose relative advantage for the production of agricultural products is improving relative to the rest of the nation, as in the Plains states for the period 1870-1910.<sup>11</sup> However, to focus our inquiry at the "1-digit" level is to conceal by aggregation the consider-

<sup>10</sup> The index of per capita agricultural production has changed as follows:

|            |        |
|------------|--------|
| 1870 . . . | 86.33  |
| 1910 . . . | 100.00 |
| 1950 . . . | 100.05 |

Source: Agricultural production data from Potter and Christy, *op. cit.*, Series AT-28; Population data from *Historical Statistics of the United States*, 1789-1945, Series B-2, and 1950 Census of Population.

<sup>10</sup> Income-elasticity of demand for food has been estimated at 0.2 to 0.3; Harlow W. Halvorson, "Long Range Domestic Demand Prospects for Food and Fiber," *Journal of Farm Economics*, Vol. XXXVI, December 1953, p. 760.

<sup>11</sup> The index of the share of the Plains states in the value of the national agri-

able variation among agricultural products at the three- and four-digit level. Some agricultural products in some specific instances can make significant contributions to regional economic growth. Thus, where cotton and cattle would rate low across the board as growth-generating items, the ability of California to engage in capital-intensive cotton production, and of Florida to exploit new breeds of cattle on the basis of excellent feed conditions, make these comparatively "good" resource products for these states. In other cases, there are agricultural specialties whose patterns of consumption have suggested a relatively high income-elasticity of demand—such as fruits, nuts, and horticultural specialties—so that they tend to contribute to the economic growth of those areas which are suited to their production.

The minerals sector has expanded much more rapidly than agriculture as a whole and regional endowment in mineral resources has always been looked upon as a positive asset for regional growth. As in the case of agriculture, there are great variations among the various mineral categories as to their contribution to regional growth. In addition, there are two characteristics of minerals that deserve special attention: first, minerals are nonrenewable resources, so that the depletion phenomenon becomes important in assessing the relative advantage conveyed by them; second, there is a high degree of substitutability among mineral products, so that the advantage of an endowment involving bituminous coal, for example, may become ephemeral as petroleum products become utilized as a substitute fuel. The impact of substitution cannot always be easily identified; it would be difficult to say, for example, how much production of steel, copper, lead, and zinc have been displaced by the growing production of aluminum. In general, the big mineral-using manufacturing industries, and particularly, the metals-using industries, have been among the most rapidly growing sectors of the economy. Also, at the level of final demand, the products of these industries have a high income-elasticity of demand. Equally important, they are the terminal products of an intricately-linked production sequence.

cultural product is

|      |        |
|------|--------|
| 1870 | 41.48  |
| 1910 | 100.00 |
| 1950 | 91.07  |

Source: Perloff, Dunn, Lampard, Muth *op. cit.* I: 38 p. 138, and I: 100 p. 249; "Regional Distribution of Value of Resources Extracted by Major Resource Industry, 1870, 1890, 1910, 1930 and 1950."

The role of petroleum and natural gas deserves a special comment. Throughout the first half of the twentieth century their production has continued to increase at a tremendous rate: over the period 1910 to 1950 the increase in output was almost elevenfold, and during this same period almost half of the total interregional shift in mining activities was accounted for by the oil-rich states of Texas, Oklahoma, and Louisiana. During the more recent period of 1939 to 1954, while employment in mining as a whole declined by 8.84 per cent, employment in this sector increased by 92.4 per cent. In terms of national levels of consumption, petroleum and natural gas have clearly been "good" resources. In terms of their multiplier effects and their linkages these resources do not rate so high. Production and refining of petroleum products is one of the most capital intensive activities in the economy so that a considerable proportion of the returns to these activities is in the form of returns to capital, which is largely imported. At the same time, while petroleum extraction and the manufacture of petroleum products are tightly linked together, the more general backward and forward linkages are relatively limited—for example, almost 80 per cent of petroleum products were destined for final demand in 1947, while absorbing only 13.1 per cent of the total inputs from other manufacturing activities. The answer to the question of how good an endowment is petroleum and natural gas is also affected by the nature of the regional supplies. These mineral fuels are strongly conditioned by the discovery-depletion cycle, so that areas narrowly specialized in the production of these mineral fuels may well find these products to have substantial disadvantages for growth if the depletion of reserves takes place at a greater rate than the augmentation of reserves by new discovery.

The fairly limited, direct, localized linkage with other economic activities is not only a characteristic of petroleum and natural gas, but of other minerals as well. For regional economic growth, the linkage between resources and other economic activities is not only a matter of *product-linkage* and value added (since the value may be added elsewhere geographically), but is even more a question of *locational linkage*—the extent to which other activities cluster in the same general area as the resources. There is some evidence to suggest that these types of "geographic" linkages are fairly limited, and that they are becoming even more so. In this category are data showing the rank correlation of employment in manufacturing with population and resources employment, by states, some of which are presented in Table 2, using data for 1954.

The proliferation of stages in the manufacturing processes has per-

TABLE 2.—RANK CORRELATIONS OF SELECTED RESOURCE-USING MANUFACTURING GROUPS WITH RESOURCE EMPLOYMENT AND POPULATION, BY STATES, 1954

| Correlated Sectors   | Coefficient |
|--|-------------|
| (1) Employment in 1st Stage Resource-Using Manufacturers with Resource Employment..... | .677        |
| (2) Employment in 2nd Stage Resource-Using Manufacturers with Resource Employment..... | .583        |
| (3) Population with Resource Employment.....   | .666        |
| (4) Employment in 1st Stage Resource-Using Manufacturers with Population.....          | .915        |
| (5) Employment in 2nd Stage Resource-Using Manufacturers with Population.....          | .935        |

NOTE: These groups were based upon input-output relationships. The industries classified as 1st stage resource users were those sectors in the 200-industry BLS table that received more than 10 per cent of their inputs (by value) from the resource sectors. The 2nd stage resource users received little directly from the resource sectors but received more than 10 per cent of their inputs from the 1st stage resource users. These two groups combined accounted for slightly less than half of the total manufacturing employment. The 10 per cent dividing line was an arbitrary choice, but was based on what seemed to be in both cases a logical division in terms of the nature of the basic productive processes involved.

SOURCE: Perloff, Dunn, Lampard, Muth, *op. cit.*, T. 148, p. 394.

mitted the increasing separation of resource processing stages from later stages. Since the processing stages are generally the primary weight-losing points in the production process, remaining stages become increasingly freed from their resource bases to seek more strategic market locations. This is reflected in the different correlations for the 1st stage and 2nd stage resource-using manufacturing groups with relation to resources employment and to population.

Employment even in the 1st stage, resource-using manufacturing industries has a high degree of geographic association with population, but a relatively limited association with resource employment (roughly equal to that between resource employment and population). The 2nd stage resource users show a higher degree of association with population and a lesser degree with resource employment. The major part of manufacturing (those not included in the two classes shown) is even further removed from resource association. For all stages of manufacturing, taken in broad categories, closeness to markets (intermediate and final) tends to be the dominant locational factor.



This underlines the point made earlier: while export of resource products provides the basis for regional economic development, extensive and continued growth can be expected to take place only in those regions which achieve sizeable regional (internal) markets. Here the notion of *cumulative* advantage is useful. Rapid advances are possible as a region reaches "threshold" size for the internal production of a wide variety of goods and services. This type of regional development is greatly enhanced where the building up of social overhead proceeds rapidly—especially in the development of an extensive internal transportation network—and where particular attention is paid to the human resources and to the conditions for living. The latter, as noted earlier, is in no small part helped by the natural conditions of the area. Where resource and nonresource advantages come together are to be found the best conditions for a high level of economic development.

#### RESOURCES IN THE RELATIONSHIP OF THE REGIONAL AND NATIONAL ECONOMIES: HEARTLAND AND HINTERLAND

In the development of the U. S. economy the role of cumulative advantage is most clearly seen in the growth of the Middle Atlantic region and, later, of the Great Lakes region. Here are regions which have enjoyed unequalled access to national market. Each was endowed with unusually good agricultural resources from the beginning, and the emergence of the minerals-dominant economy found each with excellent access to vast deposits of iron ore and coal. With these resource and market advantages, they developed into the most significant feature of regional economic growth on the American scene—the emergence of an industrial heartland coincident with the center of the national market.<sup>12</sup>

The emergence of the industrial heartland set the basic conditions for regional growth throughout the nation—it was the lever for the successive development of the newer peripheral regions: as its input requirements expanded, it reached out into the outlying areas for its resources, stimulating their growth differentially in accordance with its resource demands and the endowment of the regions. The rapid growth of the U. S. economy was accompanied (and to some extent achieved) by this process of industrial nucleation.

A major consequence of the process of expansion and regional differentiation has been the *specialization* of regional roles in the national economy, and the nature of this specialization has influenced the content

<sup>12</sup> New England, which is also part of the Manufacturing Belt, can be considered a lesser-endowed, junior partner in the industrial heartland.

and direction of regional growth. In following this process of growth, we see the working out of the general principles touched upon in the previous two sections.

Using a three-sector classification of economic activity and eight multistate regions (as in Table 1) for the period 1870-1950, we can use a simple index of specialization<sup>13</sup> to describe the dynamics of regional specialization in the national economy during this period. This is plotted for two of the three major industrial sectors—resource activities and manufacturing—in Figure 1 (100 = national average). The data serve to highlight the nucleation process. In the three regions which have coalesced into the industrial heartland—New England, Middle Atlantic, and Great Lakes—strong manufacturing specialization has characterized the entire period; however, during this eighty-year period there has been a relative decline in manufacturing specialization (i.e., as compared to the nation as a whole) in the eastern end (New England and Middle Atlantic) and a continuing increase in the western (Great Lakes) end as (1) the center of gravity of the national market shifted toward the west, and (2) the superior resource endowment of the western end helped tip the scales in its favor.<sup>14</sup> The great outlying regions (with the exception of the Far West) have maintained or increased their relative specialization in resource activities over the whole period, while in the heartland resource activities in relative terms have continuously declined, suggesting the progressive reaching out of the heartland into the hinterland areas for its resource inputs.<sup>15</sup> These data highlight the significance of the heartland-hinterland construct in the development of the national economy, and its persistence and stability in the face of dramatic structural changes during this period. The nature of this process has a number of important implications for regional growth.

$$I = 100 \times \frac{\text{per cent of region's labor force in given sector}}{\text{per cent of nation's labor force in given sector}}$$

<sup>13</sup> It is worth noting that the industrial "heartland" itself is not a static geographic area, but an area whose size and extent (and even role) shift with significant changes in the national economy.

<sup>14</sup> As in so many other indexes, the Far West (especially California) emerges as a unique case which, at least since the end of the 19th century, has followed neither heartland nor hinterland patterns, but which can be described in terms of a subnucleation in the national economy, or, if one prefers, as a second nucleation. The development of California thus suggests some interesting questions—for example, about the possibilities of "second-growth" (or new-conditions) nucleations, as well as the possibilities of a gradual spreading out of nucleation-type or high-level development in advanced stages of national economic growth. Even under the latter circumstance, the heartland-hinterland concept retains valuable explanatory power in analyzing regional development over time.



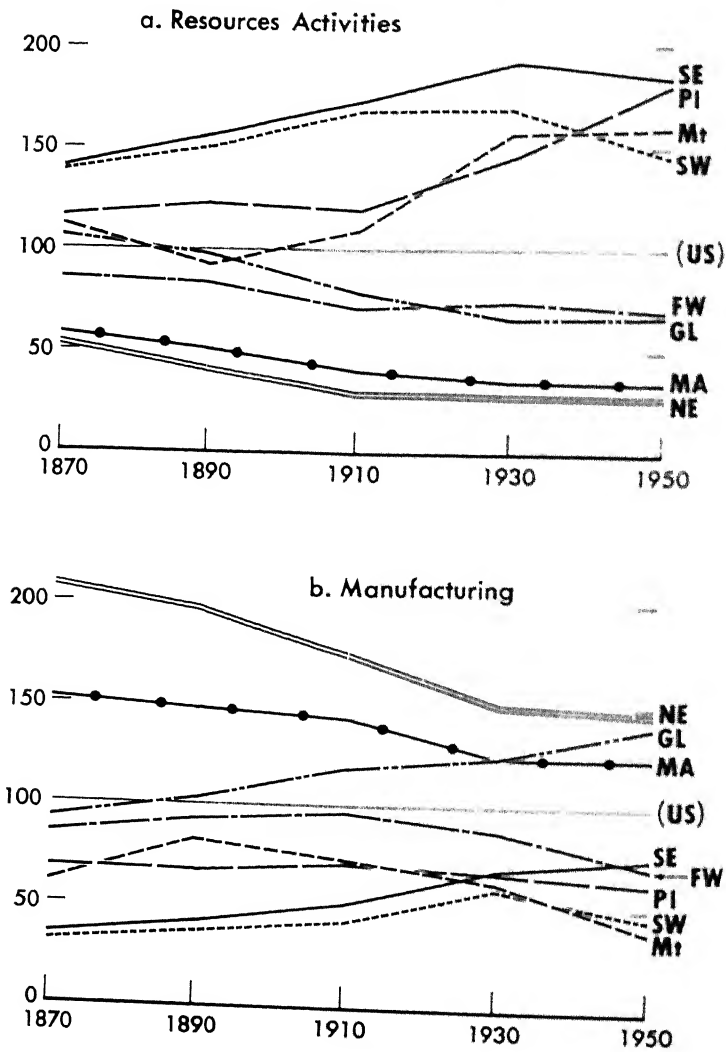


Figure 1. Indices of specialization in resources activities and manufacturing, multistate regions, 1870, 1890, 1910, 1930, and 1950 (U.S. = 100)

In the hinterland regions the working out of comparative advantage can result in a narrow and intensive specialization in a single resource subsector, in effect tying the future of the region to the vicissitudes of national demand for the products of that subsector. This will set at least ultimate limits to the region's growth rates: shifts in national demand patterns, the emergence of substitutes, depletion, technological advances, or the relative shifting of regional advantage may at any time choke off

growth and leave behind enclaves of unemployed resources and economic stagnation. At its extreme, the western experience of "boom-town to ghost-town" is a dramatic illustration, but almost as severe has been the history of the tobacco and cotton producing areas in the South. These consequences are not confined to single-product specialization. Broader, sector-wide regional specializations may produce similar problems where the degree of specialization is great and where the products in the aggregate have a low income-elasticity of demand. Typical of this kind of problem is the experience of the Plains states which has been increasing specialization in agriculture since 1910, at the same time that their relative contribution to total national value of agricultural products has been declining.<sup>16</sup>

On the other hand a broad and diverse resource specialization involving products in growing demand may provide a continuing impetus to regional expansion, especially where there is some complementarity among the resource activities. The Southwest illustrates the advantage of such a condition. Here a flourishing chemical industry has emerged based on rich endowments of petroleum, natural gas, sulfur, and salt—this is doubly fortunate, considering the high rate of growth of chemicals industries in the national economy.<sup>17</sup>

In short, the economic expansion of the hinterlands is closely associated with their resource endowments and the manner in which their endowments contribute to the evolution of favorable patterns of specialization or substantial levels of cumulative advantage.

When we look at the manufacturing sector, some important regional facets also emerge. Thus, the economic expansion of the hinterlands is accompanied by a certain amount of induced manufacturing growth. This falls into two general classes: (1) Industries devoted to the processing of regional resource products loom large. If we classify all manufacturing into first-stage resource users, or "processing" industries, and later stage, or "fabricating" industries, and plot by State Economic Areas

<sup>16</sup> See n. 12, p. 16.

<sup>17</sup> Annual growth rates of the following chemical end-products groups are suggestive:

|   | <i>Period</i> | <i>Average Annual Growth</i> |
|---|---------------|------------------------------|
| 1. Synthetic fibers (not including rayon and acetate) | 1940-1954     | 36.1%                        |
| 2. Synthetic organic plasticizers                     | 1936-1954     | 16.8%                        |
| 3. Synthetic plastics and resins                      | 1940-1954     | 16.6%                        |
| 4. Fixed nitrogen in fertilizers                      | 1939-1955     | 13.4%                        |

Source: Harold J. Barnett and Frederick T. Moore, "Long Range Growth of Chemical Industries," *Chemical and Engineering News*, April 7, 1958, p. 81.

which class of industry is dominant, we find that the processing industries dominate throughout the resource hinterlands, while the fabricating industries dominate in the industrial heartland, as shown in Figure 2. Thus, the process of industrialization not only defines the resource role of the hinterlands, but also sorts out the kinds of manufacturing activities between the heartland and the hinterland. (2) Less distinct is the role of region-serving industries in the regional growth process. These are generally market-oriented industries, producing products for regional final demand: As regions grow, their expanding markets offer increasing opportunities for economies of scale, so that one dimension of regional growth is a kind of "filling-in" generated by emerging regional market possibilities. This kind of growth frequently takes place at the expense of imports from other regions, so that one characteristic of regional growth may be a decline in the relative advantage of other regions from which imports have flowed in the past.

A general idea of this total effect is provided by a measure of the "differential shift" in manufacturing employment, as shown in Figure 3. This measures the extent to which the growth of employment in the major (two-digit) manufacturing industries within each of the states of continental United States has exceeded or fallen below "expected" growth; that is, the average national growth for each of the industry categories during a given period (here, over the period from 1939 to 1954). These "within-industry" shifts are netted out for each state, and each state total is shown as a percentage either of all the above-average growth states taken together (i.e., a percentage of the "upward shift") or of all the below-average growth states taken together (a percentage of the "downward shift"), depending on which category the state falls into. Thus, for example, Texas had 11.4 per cent of the total gain—or greater than expected increases—of all the states that experienced above-average growth in employment for all the two-digit manufacturing industries taken together.

At the same time, however, the rapid-growth (often the new) manufacturing industries have continued to find their most favorable location to be in the industrial heartland. This is shown by the "proportionality shift" in manufacturing employment—a measure of the relative change in manufacturing employment among the states due to their industrial composition.<sup>18</sup> (See Figure 4.) The significant role in regional economic

<sup>18</sup> Even if manufacturing employment in each industry had grown at the national average for the industry within each state, some states would have had a greater than average increase in total manufacturing employment because of a "favorable" industrial composition; that is, a high proportion of "rapid growth" industries. That is what is measured in the "proportionality" or composition effect.

(Note <sup>18</sup> continued on page 23)

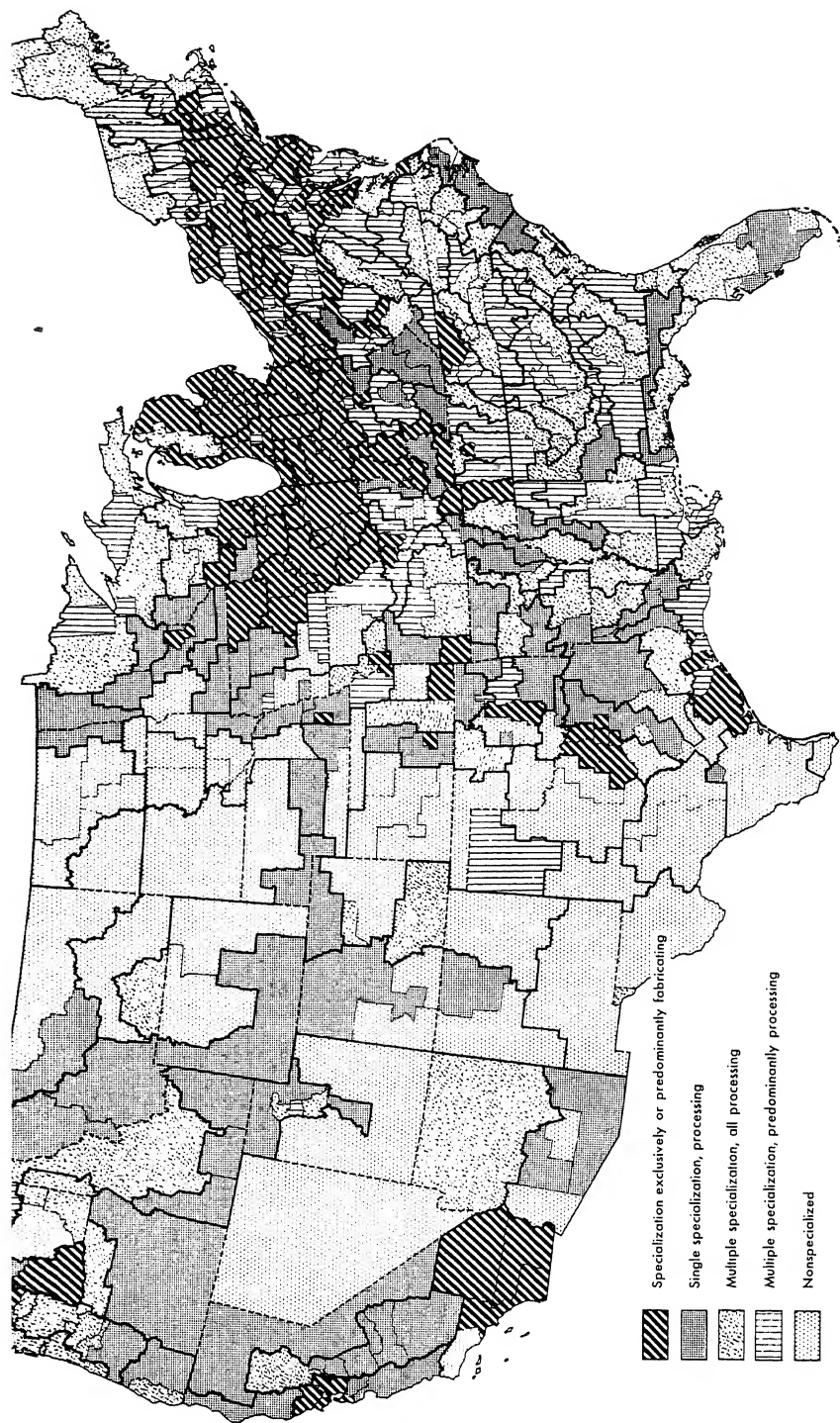
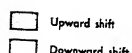
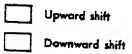


Figure 2. Type of manufacturing specialization, 1950, state economic areas.  
(Source: An unpublished series of data for state economic areas prepared by Donald J. Bogue and Calvin L. Beale.)



(Disclosure problems made it difficult to measure this dimension accurately for those states marked "d"; the direction of shift was ascertainable, however, in all states but North Dakota. The % figures are rough indications of dimension only.)



*(Disclosure problems made it difficult to measure this dimension accurately for those states marked "d"; the direction of shift was ascertainable, however. The % figures should be taken as rough indications of dimension only.)*

growth of industrial composition can be highlighted by noting the very wide range of industrial employment growth rates during the same period, 1939-54 (Table 3). The regions which have contained or attracted the machinery, metals, and chemical industries have, of course, gained volume-wise through the unusually rapid growth of these industries. It would seem that these industries find that the massive markets, economies of agglomeration, and the extensive social overhead investment of the Manufacturing Belt provide them with the economic environment most conducive to their growth and prosperity.

As highlighted by a comparison of Figures 3 and 4, the industrial heartland tends to grow in a different way than do the hinterland areas. The industrial heartland serves not only as the focal point of the national market taken as a whole, but also as the industrial seedbed of the economy. The newer products tend to be started here, nourished along, and as they find wide acceptance and volume grows, often the manufacturers find that they can supply the outlying markets more economically by producing on a decentralized basis. There are cases, of course, where the reverse is true and experimentation begins away from the center, but these have been relatively limited in number. In broad terms, the hinterland areas have grown mainly by the "filling-in" process referred to above, a threshold-by-threshold upward movement as resource exports expand and as the regional markets grow in size.

Focusing, then, on the specialized role of the regions in the national economy we get a picture in broad strokes of the spatial dimensions of the national economy. Central to it is the great heartland nucleation of industry and the national market, the focus of the large-scale, nation-serving industry, the seedbed of new industries responding to the dynamic structure of national final demand, and the center of high levels of per capita income.<sup>19</sup> Radiating out across the national landscape are the resource-dominant, regional hinterlands specializing in the production of resource and intermediate inputs for which the heartland reaches out to satisfy the input requirements of its great manufacturing

---

shown in Figure 4. For a more detailed description of these measures, see Perloff, Dunn, Lampard, and Muth, *op. cit.*, Chapter 5.

<sup>19</sup> While a discussion of the welfare aspects of regional growth is beyond the scope of this paper, it is worth noting at least briefly that among the more dramatic features of the evolving heartland-hinterland relationship has been its consequences for levels of per capita income. If we divide the nation into two parts—the heartland (including the Far West as a subnucleation) and the resource hinterland—and take the per capita personal income in the former as an index of 100, the comparable index for all the hinterland regions taken together is 58 for



TABLE 3.—PERCENTAGE CHANGE IN MANUFACTURING PRODUCTION WORKER EMPLOYMENT FOR 2-DIGIT MANUFACTURING SECTORS, 1939-1954

| S.I.C.<br>Code | Industry sector                | % Change | Employment in the<br>industry as:  |                                    |
|----------------|--------------------------------|----------|------------------------------------|------------------------------------|
|                |                                |          | % of total<br>1939 em-<br>ployment | % of total<br>1954 em-<br>ployment |
| 36             | Electrical machinery.....      | +191.39% | 3.2%                               | 5.8%                               |
| 37             | Transportation equipment.....  | +143.70  | 7.0                                | 10.7                               |
| 38             | Instruments & rel. prod.....   | +130.48  | 1.1                                | 1.6                                |
| 39             | Miscellaneous mfg.....         | +129.91  | 3.1                                | 4.5                                |
| 35             | Machinery (except elect.)..... | +118.50  | 6.9                                | 9.5                                |
| 34             | Fabricated metal prod.....     | + 82.04  | 5.8                                | 6.6                                |
| 28             | Chemical & allied prod.....    | + 81.84  | 3.5                                | 4.1                                |
| 30             | Rubber prod.....               | + 62.53  | 1.5                                | 1.6                                |
| 26             | Paper and allied prod.....     | + 61.24  | 3.5                                | 3.5                                |
|                | TOTAL MANUFACTURING EMPL....   | + 58.46  | 100.0                              | 100.0                              |
| 32             | Stone, clay & glass prod.....  | + 54.26  | 3.4                                | 3.3                                |
| 27             | Printing & publishing.....     | + 54.04  | 4.2                                | 4.0                                |
| 29             | Petroleum and coal prod.....   | + 52.11  | 1.4                                | 1.3                                |
| 25             | Furniture & fixt.....          | + 51.36  | 2.4                                | 2.3                                |
| 23             | Apparel and rel. prod.....     | + 42.10  | 9.6                                | 8.7                                |
| 20             | Food & kindred prod.....       | + 41.90  | 10.3                               | 9.2                                |
| 33             | Primary metal industries.....  | + 39.57  | 8.6                                | 7.6                                |
| 24             | Lumber & prod.....             | + 21.47  | 5.4                                | 4.6                                |
| 31             | Leather & leather prod.....    | - 1.84   | 4.2                                | 2.6                                |
| 22             | Textile mill prod.....         | - 12.41  | 13.8                               | 7.7                                |
| 21             | Tobacco manufactures.....      | - 19.78  | 1.1                                | 0.6                                |

SOURCE: Perloff, Dunn, Lampard, Muth, *op. cit.*, Table 147, p. 391.

1920 and 69 in 1950. Thus, although the per capita income differences have tended to narrow with the process of hinterland development, the advantage in welfare terms for the industrial heartland is still tremendous. As several studies have shown, the structural origin of the income differences is associated with the agricultural-nonagricultural dichotomy, so it is not surprising that we find such significant income dominance by the nuclear regions. This is, of course, not a hard and fast relationship. That a high degree of manufacturing specialization in industries with a relatively low income-elasticity of demand not only can dampen regional growth rates but depress a region's relative welfare levels has been indicated by the New England experience. At the same time, it is possible for resource regions to enjoy comparatively high per capita incomes, as the Mountain states have proven by turning up periodically with per capita incomes exceeding the national average. The important conclusion, however, is that so long as income differentials are associated with industrial structure, and structure is regionally determined by this ongoing process of nucleation, there are likely to continue to be interregional differences in welfare levels.

plant. Here in the hinterlands, resource endowment is a critical determinant of the particular cumulative advantage of the region, and hence of its growth potential. This heartland-hinterland relationship seems to be the basic morphology of the subnational economy; through it we can better understand the role of resource endowment in regional growth.

#### SUMMARY

In summary, the U. S. experience suggests the following set of propositions. "Resource endowment" is continuously redefined by changes in national final and intermediate demand, production technology, and economic organization. The relative economic growth of a region is directly related to its relative advantages in the production of goods and services for the national market; these may result from resource endowment on the one hand, or from a favorable degree of access to the national markets on the other—more generally, from a combination of the two. These advantages are normally conditioned by other elements, such as the quality of the labor supply and relative labor costs. The working out of cumulative advantage is exhibited in the specialized role that a region plays in the national economy, and this specialized role can best be described in terms of the heartland-hinterland relationships.



# 13 Location Theory and Regional Economic Growth<sup>1</sup>

Douglass C. North

## I

**D**URING the past several decades there has been a growing interest in location theory in America. Building on the pioneering works of Thünen, Weber, Lösch, Palander, and others,<sup>2</sup> a number of economists and geographers have extended the analysis to apply to a wide range of problems and have attempted to synthesize location theory with other fields of economics.<sup>3</sup> However, very little work has been done in using the principles of location to analyze the historical growth of regions in America.<sup>4</sup> While economists concerned

with location theory sometimes point out the implications of their analysis for the growth of regions, they have not followed up these discrete observations with any systematic analysis. A fundamental difficulty has been that the theory of regional economic growth<sup>5</sup> has little relevance for the development of regions in America. Not only does the sequence of stages outlined by the theory bear little resemblance to American development but its policy implications are also fundamentally misleading.

This paper will attempt to demonstrate the inadequacies of the existing theory of regional economic growth and will advance a number of propositions that may lead to a more useful theory, both for analyzing the historical development of the American economy and for understanding the contemporary problems associated with regional economic growth.

The analytical propositions advanced in this paper, though explicitly oriented to America's development, would apply equally well to other areas that meet the following conditions: (1) regions that have grown up within a framework of capitalist institutions and have therefore responded to profit maximizing opportunities, in which factors of production have been relatively mobile,<sup>6</sup> and (2) re-

<sup>1</sup> I am indebted for criticisms and suggestions to several of my colleagues at the University of Washington, particularly Philip Cartwright, J. R. Huber, Franklyn Holzman, and Robert Lampman. Dean H. W. Stoke and the Research Committee of the Graduate School at the University of Washington generously provided financial assistance for research, part of which is used in this article.

<sup>2</sup> A summary of earlier contributions to location theory may be found in E. M. Hoover, *Location Theory and the Shoe and Leather Industries* (Cambridge: Harvard University Press, 1937).

<sup>3</sup> In addition to Hoover's valuable study cited above, see his *The Location of Economic Activity* (New York: McGraw-Hill Book Co., 1948). See also Bertil Ohlin, *Interregional and International Trade* (Cambridge: Harvard University Press, 1935); National Resources Planning Board, *Industrial Location and National Resources* (Washington, D.C.: Government Printing Office, 1943); and the articles by Walter Isard cited below.

<sup>4</sup> A significant exception is Walter Isard's "Transportation Development and Building Cycles," *Quarterly Journal of Economics*, LVII (November, 1942), 90-112. See also William H. Dean, *The Theory of the Geographic Location of Economic Activities* (Selections from the Doctoral Dissertation) (Ann Arbor: Edward Bros., Inc., 1938).

<sup>5</sup> See Sec. II below.

<sup>6</sup> Obviously both profit maximization and factor mobility are relative notions and nowhere perfectly met. However, there is a vast difference between the response of an underdeveloped area where the social and economic structure is not fundamentally geared

regions that have grown up without the strictures imposed by population pressure.

## II

Both location theory and the theory of regional economic growth have described a typical sequence of stages through which regions move in the course of their development.<sup>7</sup> E. M. Hoover and Joseph Fisher, in a recent essay entitled "Research in Regional Economic Growth,"<sup>8</sup> point out that "there is now a fairly well accepted body of theory regarding the normal sequence of development stages in a region."<sup>9</sup> This sequence may be outlined as follows:

1. The first stage in the economic history of most regions is one of a self-sufficient subsistence economy in which there is little investment or trade. The basic agricultural stratum of population is simply located according to the distribution of natural resources.

2. With improvements in transport, the region develops some trade and local specialization. "A second stratum of population comes into being, carrying on

simple village industries for the farmers. Since the materials, the market, and the labor are all furnished originally by the agricultural populations, the new 'industrial superstructure' is located with reference to that 'basic stratum.'"<sup>10</sup>

3. With the increase of interregional trade, a region tends to move through a succession of agricultural crops from extensive grazing to cereal production to fruit-growing, dairy farming, and truck gardening.<sup>11</sup>

4. With increased population and diminishing returns in agriculture and other extractive industries, a region is *forced* to industrialize. "Industrialization means the introduction of so-called secondary industries (mining and manufacturing) on a considerable scale."<sup>12</sup> Typically the early stages of industrialization are based on the products of agriculture and forestry and include such activities as the processing of food, the manufacture of wood products, and the preparation of textile fibers. If industrialization is to continue, mineral and energy resources become critical.

As a second stage of industrialization, then, we see [in the regions possessing economically usable mineral resources] such industries as the smelting, refining, and processing of metals; oil refining; chemical industries based mainly

to capitalist stimuli and the kind of response one can expect in a basically capitalist society. The reluctance of the economic historian to make more extensive use of the tools of the theorist reflects in good part the fact that most of the world's economic history falls outside our first condition and that therefore economic theory is of little use in analyzing a large part of its development. On the other hand, the joint efforts of economic theorists and historians applied to the development of the United States and of some other areas hold out the promise of yielding valuable insights.

<sup>7</sup> See August Lösch, "The Nature of Economic Regions," *Southern Economic Journal*, V (July, 1938), 71-78; Hoover, *Location Theory and the Shoe and Leather Industries*, pp. 284-85, and *The Location of Economic Activity*, pp. 187-88.

<sup>8</sup> Universities-National Bureau Committee for Economic Research, *Problems in the Study of Economic Growth* (New York: National Bureau of Economic Research, 1949), chap. v.

<sup>9</sup> *Ibid.*, p. 180.

<sup>10</sup> Hoover, *Location Theory and the Shoe and Leather Industries*, p. 284. The second stage of regional growth has been elaborated by Hoover and Fisher to include some further specialization and interregional trade (*op. cit.*, p. 181).

<sup>11</sup> The theory of location diverges here from the theory of regional economic growth in stressing the historical pattern of the emergence from feudalism. Since this pattern has little meaning for American development, it is omitted here. However, it will be an important part of my argument that American location theorists have implicitly accepted a good deal of this stage sequence based on the European experience of emergence out of feudalism without recognizing the significant difference between this pattern and the pattern of American development.

<sup>12</sup> Hoover and Fisher, *op. cit.*, p. 182.

on coal, petroleum, potash, salt, and other minerals; and glass and ceramics industries. Where cheap hydroelectric power is available, industries requiring large amounts of cheap power (nonferrous metals refining, ferroalloys, and special steels, artificial abrasives, etc.) are possible, as in Norway, Switzerland, the Tennessee Valley, and the Columbia River Valley.<sup>13</sup>

5. A final stage of regional growth is reached when a region specializes in tertiary industries producing for export. Such a region exports to less-advanced regions capital, skilled personnel, and special services.

The role of transport costs has been critical in the advancement through these successive stages of growth. Isard summarizes this effect as follows: "Historically we find that reduced transport rates have tended (1) to transform a scattered, ubiquitous pattern of production into an increasingly concentrated one, and (2) to effect progressive differentiation and selection between sites with superior and inferior resources and trade routes."<sup>14</sup>

### III

When this sequence of stages is placed against the economic history of regions in America, two basic objections arise. (1) These stages bear little resemblance to the actual development of regions. Moreover, they fail to provide any insights into the causes of growth and change. A theory of regional economic growth should clearly focus on the critical factors that implement or impede development. (2) Furthermore, if we want a normative model of how regions should grow, in order to analyze the causes of

arrested development or relative decay, then this sequence of stages is of little use and is actually misleading in the emphasis it places on the need for (and difficulties of) industrialization.<sup>15</sup>

The problems of industrialization will be explored later in this paper when the causes of regional growth are examined. Here we are concerned with the first objection: the lack of correspondence between the stages of the theory of regional economic growth and the economic history of regions in America. A major discrepancy is immediately evident; namely, that America was exploited in large part as a capitalist venture. Settlement in new regions and their subsequent growth were shaped by the search for and exploitation of goods in demand on world markets. The result was a kind of development very different from that implied by the theory of regional development in which regions gradually extended the market from a subsistence economy. From the early joint-stock companies on through the whole westward expansion, a basic objective was to exploit the land and its resources in order to produce goods that could be marketed "abroad" and would bring in a money income. This is in marked contrast to the experience of Europe (which appears to be the model for the early stages of the theory of regional economic growth), where a market-oriented economy emerged only gradually from the predominantly local economies of the manorial system. If a subsistence economy existed in a new region in America, it was solely because of a lack of means of transport, a condition that was swiftly remedied by the

<sup>13</sup> Hoover, *The Location of Economic Activity*, p. 193.

<sup>14</sup> Walter Isard, "Distance Inputs and the Space Economy. Part I. The Conceptual Framework," *Quarterly Journal of Economics*, LXV (May, 1951), 188-98.

<sup>15</sup> Hoover and Fisher stress the difficulties of achieving an industrial status and maintain that most of the bottlenecks and problems of arrested development occur in moving from an agricultural to an industrial economic base (*op. cit.*, pp. 182-84).

concerted efforts of the settlers.<sup>16</sup> This is not to deny that many homesteaders maintained a subsistence existence but only to affirm that such settlement was not significant in shaping the economic development of the region, any more than the modern subsistence farmer of the backcountry is shaping the development of contemporary agriculture.

This point may be illustrated briefly from the economic history of the Pacific Northwest.<sup>17</sup> Not only has this region never experienced a subsistence economy, but its markets from the very beginning have often been thousands of miles distant. Even before general settlement, the region was exploited for its furs by the Hudson Bay Company. With the decline of the fur trade and the coming of settlers, wheat, flour, and lumber were quickly developed as exportable commodities. They first found markets in California in the 1840's. With the gold rush the demand for both wheat and lumber expanded tremendously, and the region experienced rapid growth based on these two commodities. In 1868 the first wheat shipment went from Portland to Liverpool, and by the late 1870's Pacific Northwest soft wheat had become an important part of the world wheat trade; a fleet of ships sailed from the re-

gion around Cape Horn every year. In 1857 the first shipment of flour was made to Japan, and thereafter Pacific Northwest flour found markets in Australia, Hawaii, the Orient, Europe, British Columbia, and California.<sup>18</sup> In each decade after 1850 an increasing percentage of the crop was exported either as wheat or in the form of flour. Before the end of the nineteenth century over half the crop was being exported from the region.

The history of the lumber industry reflects a similar preoccupation with markets foreign to the region. The first lumber shipment went to California in 1847, and during the gold rush lumber exports from the Pacific Northwest expanded rapidly. The rate of growth of the lumber industry was directly related to the growth of the markets reached by water (primarily California, British Columbia, and some foreign markets). In 1894 James J. Hill established a 40 cent per hundredweight freight rate on lumber to Minneapolis on his railroads, and the industry began to compete with the southern pine region for markets in the Middle West. With this rapid growth of markets the industry expanded many fold. In the first five years of the twentieth century output more than doubled, and thereafter in each successive decade Pacific Northwest fir increased its share of the national market at the expense of southern pine. The rate of growth of the region has been directly related to these basic exports. Between 1860 and 1920 lumber and flour-milling accounted for between 40 and 60 per cent of the value of the region's manufacturing output. Almost all the rest of secondary industry (as well as tertiary industry) was passive in the sense that it served local consumer needs.

<sup>18</sup> A substantial amount of the wheat and flour sent to California was exported to Europe.

<sup>16</sup> More often than not, this concerted effort was directed toward getting the government to provide the necessary internal improvements.

<sup>17</sup> This brief summary of the development of the Pacific Northwest is condensed from a larger research project that I am currently undertaking. Support for the data presented here may be found in John B. Watkins, *Wheat Exporting from the Pacific Northwest* (State College of Washington Agricultural Experiment Station Bull. 201 [May, 1926]); the Silver Anniversary Number of the *Commercial Review* (Portland, Ore.), July 1, 1915; E. S. Meany, Jr., "History of Northwest Lumbering" (Ph.D. dissertation, Harvard University, 1935); and R. W. Vinnedge, *The Pacific Northwest Lumber Industry and Its Development* (New Haven: Yale University School of Forestry, 1923).

Its growth therefore reflected the changing fortunes of the region's exportable commodities.<sup>19</sup> Wheat played a similarly critical role in the development of the region, although by the end of the nineteenth century the agricultural export base had broadened to include a number of other commodities.

This brief account of the development of the Pacific Northwest bears no resemblance to the theory of regional economic growth. There was no gradual evolution out of a subsistence economy. Instead the whole development of the region from the beginning was dependent on its success in producing exportable commodities. Nor was the Pacific Northwest's history exceptional. Furs and the products of mining were typically the early exportable commodities of western America. Colonial America exported such products as tobacco, rice, indigo, naval stores, ships, and fish. Even the well-worn historical generalization of location theorists that reduced transport rates will transform a scattered, ubiquitous pattern of production into an increasingly concentrated one is not true of America. Many new regions in America developed from the beginning around one or two exportable commodities and only widened their export base *after* transport costs had been reduced.<sup>20</sup> In short, both this generalization of location theorists and the early stages in the theory of regional economic growth appear to be taken uncritically from European experience rather than derived from the economic history of this country.

A basic starting point for reshaping our views on regional economic growth

<sup>19</sup> This point will be elaborated and qualified in the following section.

<sup>20</sup> In the Pacific Northwest the export base (particularly of agricultural commodities) only broadened after the advent of the railroad.

might well be the insights of the late Harold Innis in his studies of the growth of the Canadian economy.<sup>21</sup> Innis' early research had convinced him of the crucial importance of the export staple in shaping new economies. His subsequent studies of the growth of these staple exports were always directed toward attempting to understand "how the Canadian economy had been generated and how it had been shaped as a working economy."<sup>22</sup> An analysis of the export staples of the Canadian economy became the basis for understanding the character of that country's economic development. Moreover, it provided real insights into the political and social institutions of the country.

The term "staple" refers to the chief commodity produced by a region. It is customarily thought of as describing products of extractive industry. Since my concept of the export commodities of a region may include products of secondary or tertiary industry as well, I shall use the term "exportable commodities" (or services) to denote the individual items and the export base<sup>23</sup> to denote col-

<sup>21</sup> See *The Fur Trade in Canada* (New Haven: Yale University Press, 1920); *The Cod Fishery: The History of an International Economy* (New Haven: Yale University Press, 1940); *Problems of Staple Production in Canada* (Toronto: University of Toronto Press, 1933); and, in collaboration with A. R. M. Lower, *Settlement and the Forest and Mining Frontier* (Toronto: Macmillan Co., 1936).

<sup>22</sup> W. A. Mackintosh, "Innis on Canadian Economic Development," *Journal of Political Economy*, June, 1953, p. 188. This article provides an excellent summary of Innis' views.

<sup>23</sup> The use of the term "base" has become popular among urban land economists and city planners in the concept of the urban economic base, which refers to those activities of a metropolitan community that export goods and services to other areas. For a history of the development of the concept see Richard B. Andrews, "Mechanics of the Urban Economic Base: Historical Development of the Base Concept," *Land Economics*, XXIX (May, 1953), 161-67.

the exportable commodities (or of a region. In young regions, dependent on extractive industries, exportable commodities and staple staples are synonymous.

In new regions typically extended with a number of different ones discovering one that was initially feasible.<sup>24</sup> The success of an industry in producing an exportable commodity can be understood in terms of principles of location theory.<sup>25</sup> The extent of an exportable commodity's comparative advantage in terms of costs of production, including distributive transfer costs, tended to limit the extent of the market.

From the viewpoint of the region, the factor for the exportable commodity was an endogenous factor, but both production and transfer costs were not. Hence, new regions bent every effort to reduce these costs in their concerted efforts to promote their economic well-being. ceaseless efforts of new regions included internally subsidized internal improvements, state aid for canal construction,

experiments with silkworm culture in the colonies are a famous case in point.

For purposes it is convenient to follow the breakdown of costs into procurement, and distribution costs (see *The Location of Activity*, pp. 7-9, 15-115). While these costs reflect factor-input coefficients and, hence, procurement and distribution costs fundamentally on transfer costs.

As done a great deal of work attempting to reduce the problems of space into economic terms through the concept of distance inputs (the cost of a unit weight over a unit distance), if a distance input is the transport rate in the case of capital inputs, a reduction in both a scale and a substitution effect. Inputs are conceived to be simply another reduction whose price is the transport cost whose optimum combination with other inputs is determined by the principles of substitution. his "Distance Inputs and the Space Problem" (p. 116).

tion, federal and state aid for railroads, and river and harbor improvements were a part of the continuous effort of each region to reduce transfer costs to better the competitive position of its exports.<sup>26</sup>

As regions grew up around the export base, external economies developed which improved the competitive cost position of the exportable commodities. The development of specialized marketing organization, improved credit and transport facilities, a trained labor force, and complementary industries was oriented to the export base.

The concerted effort to improve the technology of production has been equally important. Agricultural experiment stations, state universities, and other local research groups become service adjuncts to export industries and conduct research in technological improvements in agriculture, mining, or whatever manufacturing comprises the region's export base.

The purpose of this concerted effort is better to enable the region to compete with other regions or foreign countries for markets. In new regions highly dependent on extractive industry, these external economies and technological developments tend to more than counteract diminishing returns in the staple product.<sup>27</sup> As a result, these efforts tend to reinforce a region's dependence on its existing staples rather than promote changes in the export base. This conservative bias is further reinforced by the

<sup>24</sup> Such efforts have not been confined to pressure-group activity but have erupted into political movements. The Grangers and the Populists were fundamentally concerned with a number of economic measures that would, for example, improve the position of American wheat on the world wheat market or provide the western miner with a better market for his silver.

<sup>27</sup> In the case of mining this statement probably would not hold.



role of capital. Capital is typically imported into new regions in the development of the export staple industries. Indeed, until a region develops sufficient income to provide a substantial share of its own investment capital, it must rely upon outside sources. External suppliers of capital tend to invest primarily in existing export industry rather than in new, untried enterprises.<sup>28</sup>

#### IV

The following section will deal with the way in which regions grow; first, however, we must explore the significance of the export base in shaping the whole character of a region's economy.

At the outset, export industries must be clearly distinguished from "residential industries."<sup>29</sup> The term "residential" is used to designate industry for the local market which develops where the consuming population resides. In order to determine the market area of each industry more precisely than can be done by a priori classification, the "location quotient" developed by Hildebrand and Mace<sup>30</sup> is employed. The location quotient measures the concentration of employment in a given industry in one area (the "subject economy," which for our purposes is the region) with another area

(the "benchmark economy," which for our purposes is the nation).

Formally the location quotient is the numerical equivalent of a fraction whose numerator is employment in a given industry in the subject economy relative to total employment in the subject economy and whose denominator is employment in the given industry in the benchmark economy relative to total employment in the benchmark economy. *A priori* a location of 1.00 means no greater relative specialization in the subject economy than in the benchmark economy, for the particular industry. In each industry values significantly below 1.00 indicate much greater relative specialization in the benchmark economy; or if well over 1.00 much greater relative specialization in the subject economy.<sup>31</sup>

*Review of Economics and Statistics*, XXXII (August, 1950), 341-49.

P. Sargent Florence developed the concept of a coefficient of localization. He first computed a "location factor" for each industry by computing the ratio of the percentage of employment in the given region found in the given industry to the corresponding percentage for the nation as a whole. If all industries were perfectly evenly distributed among regions, the location factor would be unity. "The coefficient of localization for a given industry is obtained by computing the weighted average deviation from unity of the location factors for all regions, the weight for a local region being the proportion of total national employment found in that region. This measure divided by two varies between zero and unity" (Vining, *op. cit.*, pp. 40-51). Completely even geographic distribution would give a coefficient of zero, while increasingly greater concentration of industry in a region would give a coefficient approaching unity. Although this method is somewhat different from that of Hildebrand and Mace, the result is the same.

<sup>31</sup> Hildebrand and Mace, *op. cit.*, p. 243. In their study of Los Angeles County these authors varied the subject and benchmark economies. Using the United States as the benchmark economy, they used successively the twelve western states, the eleven counties of southern California, and Los Angeles County as subject economies. Then using the eleven western states as the benchmark economy, they used southern California and Los Angeles County as subject economies and finally used Los Angeles County relative to southern California. As a result, they were able to delimit precisely the extent of the market for each export (while exports out of the country would increase the location quo-

<sup>28</sup> This outside capital often comes in waves associated with (or in anticipation of) substantial reductions in costs or increases in demand. As a result the growth of regions tends to be uneven. This whole subject of the growth of regions is dealt with in more detail in Sec. V.

<sup>29</sup> The term "residential industry" was first used by P. Sargent Florence in National Resources Planning Board mimeographed releases. Rutledge Vining subsequently employed the concept in "Location of Industry and Regional Patterns of Business Cycle Behavior," *Econometrica*, XIV (January, 1946), 37-68.

<sup>30</sup> George Hildebrand and Arthur Mace, Jr., "The Employment Multiplier in an Expanding Industrial Market: Los Angeles County, 1940-47,"

Thus industries producing for export will show values significantly above 1.00.<sup>32</sup>

We are now in a better position to examine the role of the export base in shaping the economy of the region.

Clearly the export base plays a vital role in determining the level of absolute and per capita income of a region. While the return to factors of production<sup>33</sup> in the export industries indicates the direct importance of these industries for the well-being of the region, it is the indirect effect that is most important. Since residentiary industry depends entirely on demand within the region, it has historically been dependent on the fate of the export base.<sup>34</sup> Vining's analysis indicates

that employment in residentiary industry tends to bear a direct relationship to employment in export industries. The median figure for employment in residentiary industry in individual states was approximately 55 per cent of the total employment.<sup>35</sup>

The export staple plays an equally vital role in the cyclical sensitivity of the region; it acts as the "carrier" in diffusing changes in the level of income from other regions to the subject region. Furthermore, the sensitivity of the region to fluctuations depends on the income elasticities of the export staples. Clearly regions that specialize in a few products with high income elasticities will have more violent fluctuations in income than more diversified regions.<sup>36</sup>

When we turn to the role of exports in shaping the pattern of urbanization and nodal centers,<sup>37</sup> we are on ground that has been more thoroughly explored by location theorists and geographers.<sup>38</sup> Again, however, the pioneering work has been done by German location theorists who have extended the implications of each stage of economic growth to embrace the logical pattern of urbanization that would ensue.<sup>39</sup> Since these

<sup>32</sup> Hildebrand and Mace allowed for differences in demand functions, which might make some residentiary industry appear with a location quotient above 1.00. They came to the conclusion that 1.508 was the boundary line in their study (*ibid.*, p. 246).

This location quotient is not too well adapted to use in agriculture. There I have used a coefficient of specialization in which the numerator is the region's physical volume of production relative to the physical volume of production of the agricultural good for the nation. The denominator is the region's absolute population relative to the nation's absolute population. While such a coefficient has some obvious limitations and must be used with care, it is more adaptable to the available data than the one discussed above.

<sup>33</sup> Obviously the disposition of nonwage income to residents of the region or outside the region is important here. It will be further considered in the next section.

<sup>34</sup> This statement requires both substantiation and careful qualification. This article is primarily concerned with the historical development of the American economy, and here the statement needs little qualification. The fortunes of regions have been closely tied to their export base. However, it is conceivable that a region with a large influx of population and capital might simply "feed upon itself" and thereby account for a substantial share of its growth. Moreover, in older "mature" regions, economic activity may become so diversified as to make the export base less significant. This question will be dealt with in the next section.

<sup>35</sup> Vining, *op. cit.*, p. 49.

<sup>36</sup> For further discussion of this subject see Vining, *op. cit.*

<sup>37</sup> The concept of nodes is one that has been extensively used by geographers. The term refers to *sites* that have strategic transfer advantages in reference to procurement and distribution costs and therefore become processing centers. Such advantageous points are limited in number and tend to develop into major metropolitan areas. For further discussion of nodes see Hoover, *The Location of Economic Activity*, pp. 119-30.

<sup>38</sup> For a summary of recent developments in this area see Walter Isard, "Current Development in Regional Analysis," *Weltwirtschaftliches Archiv*, LXIX (September, 1952), 81-91.

<sup>39</sup> An excellent summary of the German contributions is contained in Isard's "The General



stages do not fit the American development, the pattern of American urbanization likewise differs in many respects from the German models. However, it is beyond the scope of this article to explore the whole question of urbanization and the export base. We may note in passing the observations of August Lösch that in such areas as Iowa, with a rather even distribution of production of agricultural staples, the distances between towns increase with their size.<sup>40</sup> In contrast, cities in the English coal districts are the same distance from each other irrespective of size.<sup>41</sup>

While discussion of the spatial distribution of urban areas would take us too far afield, the role of the export base in shaping the growth of nodal centers deserves some attention. Nodes grow up because of special locational advantages that lower the transfer and processing costs of exportable commodities. Nodal centers become trading centers through which exports leave the region and imports enter for distribution throughout the area. Here special facilities develop to implement the production and distribution of the staples. Subsidiary industries to service the export industry, as well as specialized banking, brokerage, wholesaling, and other business services, concentrate in these centers and act to improve the cost position of the export.<sup>42</sup>

The character of the labor force will be fundamentally influenced by the export industries. The types of skills required,

the seasonality and stability of employment, and the conditions of work will shape the social attitudes of the working force.

As already noted, the political attitudes of the region will be largely directed toward improving the position of its export base. The extent of such activity is too well known historically and too obvious in the contemporary American political scene to require extended discussion.

## V

Previous sections of this paper have examined the significance of the export base for a region's economy. I have tried to indicate the primary role that such exports have played historically, but I have not yet touched on the critical question of the causes of the growth of a region. It is evident that this growth is closely tied to the success of its exports and may take place either as a result of the improved position of existing exports relative to competing areas or as a result of the development of new exports. However, a major question that must first be examined is whether a region must industrialize if it is to continue to grow. Such a necessity has been a major tenet of the theory of regional economic growth. Moreover, industrialization has been regarded as a difficult stage to achieve, so that it is the source of problems of arrested regional development. Hoover and Fisher stress three factors that make this transition difficult: (1) the need for greatly improved transportation facilities, which call for large-scale capital in-

Theory of Location and Space Economy," *Quarterly Journal of Economics*, LXIII (November, 1949), 476-506.

<sup>40</sup> Lösch, *op. cit.*, p. 75. In this article Lösch advances an interesting theoretical model of spatial location.

<sup>41</sup> *Ibid.*, p. 75. A summary of the development of concepts of spatial organization may be found in Isard's "Distance Inputs and the Space Economy," *op. cit.*

<sup>42</sup> These specialized facilities provide economies in addition to the general economies of urban concentration resulting from such things as fire and police protection, lower utility rates, and a specialized labor force. For further discussion of these aspects of urban concentration see Ohlin, *op. cit.*, pp. 203-4.

vestments; (2) the need for intensification of the geographic division of labor; and (3) the fact that industrial technology is novel to an agricultural region.<sup>43</sup> If these statements are correct, then the implications for our analysis are clear. At some point regions must shift from an extractive to an industrial export base, and this shift will be fraught with difficulties. However, the contention that regions must industrialize in order to continue to grow, as well as the contention that the development of secondary and tertiary industry is somehow difficult to achieve, are both based on some fundamental misconceptions.

The importance of industrializing is based upon the notion that, with increased population and diminishing returns in extractive industry, the shift to manufacturing is the only way to maintain sustained growth (measured in terms of increasing per capita income). This argument has been buttressed by evidence such as that gathered by Dr. Louis Bean correlating per capita income with percentage of the labor force engaged in primary, secondary, and tertiary occupations by states for 1939.<sup>44</sup> Bean's figures purport to demonstrate that increased industrialization leads to higher per capita income, and he goes so far as to say that "a 10-point [per cent] increase in industrial progress in the east and south . . . apparently tends to add \$100 to \$150 (1939 prices) per capita and in the western states substantially more."<sup>45</sup> In fact, Bean's statistics do not

prove this, and the policy implications of such generalizations may be misleading and dangerous.

We may note first of all that his correlation is not very impressive. There were eleven states in which the percentage of the labor force in primary occupations was above the national average whose per capita income either exceeded the national average or was close enough to the average so that annual variations could well place it on one side or the other. Indeed, had the correlation been made for postwar years, it would have been substantially different.<sup>46</sup>

Furthermore, money-income data significantly understate the real income of the farmer,<sup>47</sup> because of the great variety of goods and services produced on the farm that require cash payment in the city.<sup>48</sup>

However, the real source of error has resulted from a basic misunderstanding of the nature of the economy. A state whose export base consists mostly of agricultural products may have a low percentage of its labor force in primary activity and a high percentage in tertiary occupations and yet be basically dependent upon agriculture for the high per capita income it enjoys. It is the agricultural export staples that provide the high income that enables the state to support a substantial level of services. In such a case both the secondary and the tertiary activities are "residential" and can sur-

<sup>46</sup> See "State Income Payments in 1950," *Survey of Current Business*, August, 1951, p. 18.

<sup>47</sup> There is also evidence to indicate that money incomes are disproportionately understated.

<sup>48</sup> See Margaret Reid, "Distribution of Non-money Income," *Studies in Income and Wealth*, Vol. XIII (New York: National Bureau of Economic Research, 1951). See also Jacob Viner, *International Trade and Economic Development* (Glencoe, Ill.: Free Press, 1952), pp. 63-73. Professor Viner provides a number of trenchant criticisms of Bean's argument.

<sup>43</sup> *Op. cit.*, p. 182. Hoover and Fisher go on to point out that "further difficulty arises from the fact that when a non-industrial region reaches a limit of growth it is likely to retrogress or decay" (*ibid.*, p. 184).

<sup>44</sup> *Studies in Income and Wealth*, VIII (New York: National Bureau of Economic Research, 1946), 128-29.

<sup>45</sup> *Ibid.*, p. 137.

vive only because of the success of the basic agricultural export staples. In short, a percentage shift in such a state from primary to secondary and tertiary employment does not necessarily reflect a shift away from dependence on agriculture to dependence on manufacturing and services. Instead, it may reflect the simple fact that farmers are receiving high incomes for their staple crops and therefore buy more goods and services from residentiary industry.

This brings us to the related question of the difficulty of industrialization. The implication of the preceding paragraph is that a substantial amount of secondary industry of the residentiary variety will develop automatically as a result of high incomes received from the exportable commodities. Nor is this the only kind of manufacturing that can be expected to develop. We may distinguish four different kinds of manufacturing that will develop:<sup>49</sup>

1. Materials-oriented industries which, because of marked transfer advantages of the manufactured product over the raw material, locate at the source of the latter. Among the industries in this category are sugar-beet refining, flour-milling,<sup>50</sup> and lumbering.<sup>51</sup> Such industries may develop further stages of vertical integration until transfer cost advantages become equalized. Such industry is typically part of the export base.

2. Service industries to the export industry. Foundries and establishments that make machine tools, specialized

agricultural implements, and logging and lumbering equipment are illustrations.

3. Residentiary industry producing for local consumption.

4. Footloose industries, where transfer costs are not of significant importance in location. A great many such industries develop purely by chance in some location.<sup>52</sup>

While footloose industries have typically developed by chance, the other three types of secondary activity develop naturally because of locational advantages in a society responsive to profit-maximizing stimuli. There is nothing difficult about the development of such industries. The difficulties arise when promoters seek to develop in a region industries which simply are unsuited for the area and which can therefore only be maintained under hothouse conditions.<sup>53</sup>

The argument may be advanced that the kinds of industry described above do not constitute industrialization. How much and what kind of secondary industry must a region possess to be termed "industrialized"? By 1950 census classification, the state of Oregon had almost

<sup>49</sup> For further discussion of such industries see National Resources Committee, *The Structure of the American Economy*, Part I: *Basic Characteristics* (Washington, D.C.: Government Printing Office, 1939), p. 36.

<sup>50</sup> This does not mean that there is no room for appropriate public policy that may create the social overhead benefits that will make certain industries feasible. I can do no better here than to quote Viner: "There are no inherent advantages of manufacturing over agriculture, or, for that matter, of agriculture over manufacturing. It is only arbitrarily in fact that the line separating the two can be drawn. The choice between expansion of agriculture and expansion of manufactures can, for the most part, be left to the free decisions of capitalists, entrepreneurs and workers. To the extent that there is need for government decision, it should be made on rational grounds, in the light of considerations of costs and comparative returns from alternative allocation of scarce national resources, human and material" (*op. cit.*, p. 72).

<sup>49</sup> This classification is similar to that of E. J. Cohn, Jr., in *Industry in the Pacific Northwest and the Location Theory* (New York: Columbia University Press, 1954), pp. 42-44.

<sup>50</sup> However, milling-in-transit privileges may modify this materials orientation.

<sup>51</sup> See National Resources Planning Board, *op. cit.*, chap. vi, for a further account of such industries.

24 per cent of its labor force in manufacturing, which was only slightly under the United States average (25.9 per cent) and exceeded the United States average in durable goods (16.7 per cent as compared with the national average of 13.8 per cent). It was well ahead of the neighboring states of Washington and California, despite the fact that these two states had a variety of manufacturing industries, in contrast to Oregon's specialized dependence on the Douglas fir lumber industry. Is such a state industrialized? Implicit in the concept appears to be the notion that industrialization is somehow tied up with steel and the capital goods industries. However, historically, the locational pull of coal and iron ore has shaped the development of the steel-producing centers, which in turn have attracted and concentrated heavy industry.<sup>54</sup> While locational influences in the steel industry have been changing significantly in the last half-century with the growing importance of scrap and the changing composition of inputs,<sup>55</sup> nevertheless, the possible areas for the development of efficient large-scale steel production<sup>56</sup> and, therefore, capital goods industry are severely circumscribed. A more useful concept of industrialization for our purposes is a region whose export base consists primarily of finished consumers' goods and/or finished manufactured producers' goods.

We may summarize the argument up to this point as follows: (1) There is no

reason why all regions must industrialize in order to continue to grow. (2) A great deal of secondary (and tertiary) industry will develop automatically either because of locational advantages of materials-oriented industry or as a passive reflection of growing income in the region resulting from the success of its exportable commodities. (3) The concept of industrialization is an ambiguous one that needs further clarification if it is to be useful.

Since the growth of a region is tied to the success of its export base, we must examine in more detail the reasons for the growth, decline, and change in the export base. Clearly, the decline of one exportable commodity must be accompanied by the growth of others, or a region will be left "stranded."<sup>57</sup> Among the major reasons<sup>58</sup> for the decline of an existing exportable commodity have been changes in demand outside the region,<sup>59</sup> exhaustion of a natural resource,<sup>60</sup> increasing costs of land or labor relative to those of a competing region,<sup>61</sup> and technological changes that changed the relative composition of inputs.<sup>62</sup> A historically important reason for the growth of new exports has been major developments in transport (in contrast with mere cost-reducing improvements in transport, which may reinforce dependence on existing exports). Such developments

<sup>57</sup> The cut-over region in the Great Lakes area is a case in point.

<sup>58</sup> For further discussion on shifting industry see National Resources Planning Board, *op. cit.*, pp. 92-104.

<sup>59</sup> Such as the decline in the demand for beaver hats, which affected the fur trade.

<sup>60</sup> Exemplified by the Great Lakes lumber industry.

<sup>61</sup> The most famous example is the decline in the New England cotton textile industry.

<sup>62</sup> Such as the case of steel cited above.

<sup>54</sup> National Resources Planning Board, *op. cit.*, p. 162.

<sup>55</sup> Walter Isard, "Some Locational Factors in the Iron and Steel Industry since the Early Nineteenth Century," *Journal of Political Economy*, LVI (1948), 213-17.

<sup>56</sup> The extensive utilization of scrap makes possible small-scale steel production as a residuary industry wherever the local market achieves sufficient size.

have often enabled a region to compete with other regions in the production of goods that were previously economically unfeasible because of the high transfer costs.<sup>63</sup> Growth in income and demand in other regions<sup>64</sup> and technological developments<sup>65</sup> have also been important. The role of the state and federal government in creating social overhead benefits has created new exports in many regions,<sup>66</sup> and the significance of war in promoting industries that may either continue or leave a residue of capital investment for peacetime use has also been important.

A region may expand as a result of increased demand for its existing exportable commodities, whether due to an increase in the income of the market area or to a change in taste. An improvement in the processing- or transfer-cost position of the region's staples vis-à-vis competing regions will likewise promote growth.

Historically, in a young region, the creation of a new export or the expansion of an existing export has resulted in the influx of capital investment both in the export industry and in all the kinds of passive and supporting economic activity described above. Meier has described this process for the Canadian economy in the first decade of the twentieth century, when increased world demand for wheat not only led to an expansion of ware-

housing, transport, public utilities, and construction in the Prairie Provinces, but also, by increasing income, augmented demand for secondary products and thereby induced investment in a host of other industries.<sup>67</sup> As a result the growth of a region will, in all likelihood, be uneven, coming in spurts of increased investment rather than proceeding at an even pace.

Increased capital investment in the export industry will go toward achieving optimum size of the enterprise, increased mechanization of the processes, and further development of the specialized services to the export. The source of capital will play an important part in the region's growth. Typically, the capital in young regions comes from outside. Profits (and some other nonwage income) flow out of the region. To the extent that the export base is profitable, a part of this income will be reinvested in the expansion of this base.

With the growth of population and income, indigenous savings will increase. Both indigenous savings and the reinvested capital can pour back into the export industries only up to a point, and then the accumulated capital will tend to overflow into other activity. As described above, some will go into residential industry and industries subsidiary to the export; but it is also very likely that some will go into locationally "footloose" industries, which may start out to serve only the region, but which can expand into export industries.

<sup>63</sup> The whole history of canal and railroad development contains innumerable illustrations of such developments (see Isard, "Transportation Development and Building Cycles," *op. cit.*).

<sup>64</sup> The growth in demand for wheat in England and on the European continent in the last half of the nineteenth century is a famous example.

<sup>65</sup> The development of the petroleum industry is a typical illustration.

<sup>66</sup> The development of hydroelectric power in the Pacific Northwest and the resultant development of the aluminum industry is an example.

<sup>67</sup> G. M. Meier, "Economic Development and the Transfer Mechanism," *Canadian Journal of Economics and Political Science*, XIX (February, 1953), 1-19. M. C. Daly has attempted to work out a geographic multiplier between "localized" and "non-localized" industry, using data for Britain for the years 1921-31 ("An Approximation to a Geographic Multiplier," *Economic Journal*, L [June-September, 1940]), 248-58. See also Hildebrand and Mace, *op. cit.*

At this point a region is no longer young. The social overhead benefits that have been created through political pressure or as a part of the pattern of urban development and the development of a trained labor force and indigenous capital make it far easier to develop new exports. Whether such industries were originally residentiary and, by gradually overcoming transfer-cost disadvantages, became export industries, or were originally footloose industries not significantly affected by transfer costs, the result is to broaden the export base. As such a region matures, the staple base will become less distinguishable, since its production will be so varied.

We may expect, therefore, that the differences between regions will become less marked, that secondary industry will tend to be more equalized, and indeed in economic terms that regionalism will tend to disappear.

## VI

The purpose of this paper has been to re-examine location theory and the theory of regional economic growth in the light of the historical development of regions in America and to advance some propositions that may lead to a new theory of regional economic growth.

It has been argued that the stages outlined in the theory of regional economic growth bear little relationship to the character of American development and more specifically do not focus on the crucial elements that will enable us to understand that growth. Furthermore, the traditional theory has policy implications that may be fundamentally in error.

The first stage of subsistence has been relatively unimportant, and, to the extent that it existed at all, it was because

means of transport were lacking rather than because of a nonmarket orientation. In Europe a subsistence or a village economy with local markets was built into the social and economic structure for centuries. In America subsistence was only a frontier condition to be overcome as rapidly as means of transport could be built.

The second stage of the theory is based on a gradual widening of the market area with improved transport and the development of a second stratum to service the basic agricultural stratum. Far from moving through such a gradual progression American regions, as soon as any transport permitted, developed goods for export often to markets thousands of miles away. The early town centers were located not only so as to service the agricultural stratum but so as to implement the export of the region's staples. The prosperity of the region depended on its success in competing with other areas producing the same staple exports. Therefore, the region's economic and political efforts were oriented toward the reduction of processing and transfer costs. The struggle for internal improvements by the West, the agrarian pressure for inflation and cheaper credit, and the campaign for free coinage of silver were fundamentally economic movements. Their objectives included increasing the supply of capital, eliminating real or fancied transport discrimination, reducing interest rates, and improving the market for silver, however much they may also have been concerned with social justice.

The third stage of regional growth has been described as the gradual shift from extensive to intensive farming. While it is true that rising land values promoted such a shift, there were many other reasons for a shift in the staple base. New

means of transport, changing demand, new technological developments, changing cost relationships vis-à-vis competing regions, government subsidization of social overhead benefits, and war have all been important.

The shift from an agricultural to an industrial base has been looked upon as the difficult, but indispensable, step for sustained economic growth. It is a major argument of this paper that such a step may be neither necessary nor desirable and that the evidence customarily advanced to support this argument proves nothing of the sort. There is nothing to prevent population and per capita income from growing in a region whose export base is agricultural. Moreover, there is nothing difficult about developing secondary and tertiary industry in such a region. Indeed, it will develop automatically, often to such an extent that analysis of the region in terms of distribution of employment will lead to the conclusion that it is an industrial region.

The final stage has typically been conceived to be the mature regional economy exporting capital, skills, and specialized services to less-well-developed regions. While this may be true for some regions, it is unlikely to be a final stage for all. Indeed, one would presume that some sort of balanced relationship would emerge among regions as transfer costs become less significant and income differentials tend to be ironed out by long-run factor mobility.

The major propositions that emerge from this paper are:

1. For economists' purposes the concept of a region should be redefined to point out that the unifying cohesion to a region, over and beyond geographic similarities, is its development around a common export base. It is this that makes it

economically unified and ties the fortunes of the area together. This tends to result in the interdependent development within the region of external economies and unified political efforts for government assistance or political reform. The geographer has emphasized the distributive functions of the nodal centers of a region, but the role of the nodal center in providing external economies for the export industries has been equally important.

2. The success of the export base has been the determining factor in the rate of growth of regions. Therefore, in order to understand this growth, we must examine the locational factors that have enabled the staples to develop.

3. The importance of the export base is a result of its primary role in determining the level of absolute and per capita income in a region, and therefore in determining the amount of residentiary secondary and tertiary activity that will develop. The export base has also significantly influenced the character of subsidiary industry, the distribution of population and pattern of urbanization, the character of the labor force, the social and political attitudes of the region, and its sensitivity to fluctuations of income and employment.

4. In a young region dependence on staples is reinforced by the concerted efforts of the region's residents to reduce processing and transfer costs through technological research, and state and federal government subsidization of social overhead benefits, as well as the tendency for outside suppliers of capital to reinvest in the existing staple base.

5. Some regions, because of locational advantages, have developed an export base of manufactured products, but this is not a necessary stage for the sustained



growth of all regions. A great deal of secondary and tertiary industry will result from the success of the export base. This residentiary industry will, in all likelihood, provide for widening the export base as a region develops.

6. The growth of regions has tended to be uneven. A given increase in demand for the region's exports (or a significant reduction in processing or transfer costs) has resulted in a multiple effect on the region, inducing increased investment not only in the export industry but in all other kinds of economic activity as well.

7. As a region's income grows, indigenous savings will tend to spill over into new kinds of activities. At first, these activities satisfy local demand, but ultimately some of them will become export industries. This movement is reinforced by the tendency for transfer costs to become less significant. As a result, the export bases of regions tend to become more diversified, and they tend to lose their identity as regions. Ultimately, we may expect with long-run factor mobility more equalization of per capita income and a wider dispersion of production.



# 14 Exports and Regional Economic Growth

Charles M. Tiebout

## I

THE theory of the regional economic base has been bobbing around in the literature, implicitly and explicitly, for some time.<sup>1</sup> Its latest appearance comes as an explanatory factor in regional economic growth. In his recent article Douglass C. North has suggested that the theory of regional development which sees the region as passing through various stages—primary, secondary, and tertiary—is not adequate.<sup>2</sup> As a substitute, North maintains that a region's growth "is closely tied to the success of its exports and may take place either as a result of the improved position of existing exports relative to competing areas or as a result of the development of new exports."<sup>3</sup> He further points out that it is necessary to look into location theory to explain changes in the export base. The point involved is that the concept of the export base in regional analysis is called on as the major autonomous variable determining the level of regional income.

The concept of the economic base has been developed largely in the works of city planners and other researchers interested in urban problems.<sup>4</sup> As such—and this is neither slur nor praise—no attempt has been made to relate this concept to the general theory of income determination as used in

national income analysis.<sup>5</sup> This failure and the continual identification of the exports of a region with the autonomous variable determining income have led to some erroneous conclusions about regional income and regional development. The purpose of this note is, first, to show how the export-base concept fits within the more general theory of income determination and, second, using this setting as a frame of reference, to point out some implications for the theory of regional growth.

## II

It is useful to begin by presenting a simplified version of the concept of the export base. The economic activities of a region are divided into those which produce for the export market and those which produce for the local market. In defining exports allowance is made for such items as the earnings of commuters, capital flows, government transfers, and linked industries. Given these basic or export activities, the level of non-basic or residuary activities follows. The ratio between export activities and residuary activities, measured in income or employment, is then used as a multiplier. For example, a one-to-one ratio would mean that an increase in exports will cause an equal increase in residuary activities. Whether or not this function is constant at all levels of in-

<sup>1</sup> See Richard B. Andrews, "The Mechanics of the Urban Economic Base," *Land Economics*, Vol. XXIX (1953), No. 3 (continuing series). For a more explicit statement see George Hildebrand and Arthur Mace, Jr., "The Employment Multiplier in an Expanding Industrial Market: Los Angeles County, 1940-47," *Review of Economics and Statistics*, XXXII, No. 3 (August, 1950), 241-49.

<sup>2</sup> "Location Theory and Regional Economic Growth," *Journal of Political Economy*, LXIII (June, 1955), 243-58.

<sup>3</sup> *Ibid.*, p. 251.

<sup>4</sup> See Richard Andrews, "The Mechanics of the Urban Economic Base: Historical Development of the Base Concept," *Land Economics*, XXIX (August, 1953), 161-67; and Homer Hoyt, "Homer Hoyt on the Concept of the Economic Base," *Land Economics*, XXX (May, 1954), 182-86.

<sup>5</sup> It is interesting to note that the work of Hildebrand and Mace (*op. cit.*), which deals with an employment multiplier, is rarely mentioned in discussions of the economic base. North's article is a notable exception.

come is not stated. There is no a priori reason to believe it is. From here, of course, it is a simple step to the statement that the income of the region is tied to the level of exports. For a small region this may be substantially correct, but for larger regions it is an oversimplification. A general theory of income determination at the national level rests on a knowledge of the level and stability of both the dependent and the autonomous variables. These are the necessary ingredients of an econometric model that forecasts income.<sup>6</sup>

There is no reason to assume that exports are the sole or even the most important autonomous variable determining regional income. Such other items as business investment, government expenditures, and the volume of residential construction may be just as autonomous with respect to regional income as are exports.<sup>7</sup> Under the assumption, which may have some validity, that the autonomous variables are the dynamic factors in determining the short-run level of regional income, these items may even be the chief source of instability. Only empirical studies will enable us to say something about their quantitative importance.

A further consideration will help to point up the error of identifying exports as the sole source of regional income change. In an exchange economy one person considered in a spatial context may be entirely dependent on his ability to export his services. Probably this is true of a neighborhood area, except for the corner grocer. For the community as a whole, the income originating in non-exports increases. In the United States economy, exports account for only a small part of national income. Obviously, for the world as a whole, there are no exports.

Thus the quantitative importance of exports as an explanatory factor in regional income determination depends, in part, on the size of the region under study. It is true

that for a region considered at two different time periods, a change in the volume of exports may indicate a change in the level of income, but this is not enough. A region may grow with exports at a constant level, if internal autonomous activities are on the upswing. The larger the region, the more the dynamic forces causing income change will be found inside its borders.

The problem that arises because export volume is a function of regional size might be solved if it were possible to find some method of determining the boundaries of a region which not only made sense but allowed for interregional comparisons. North has suggested that the boundaries of a region should be determined by "its development around a common export base."<sup>8</sup> This basis of classification is useful, but it is by no means the only possibility.

Most researchers in the field of regional economics have come to the conclusion that there is no "ideal" region. Probably the closest approximation to the concept of an ideal region would arise in a Lösch production-oriented spatial system.<sup>9</sup> In this system an over-all area is mapped out according to sites of production determined by market networks. Other conditions which are also given for equilibrium need not concern us here. In the central city all goods are produced, with fewer produced in the other spatially arranged cities. If an over-all area, in this sort of orientation, could be divided into two or more identical parts, either one might be considered an ideal region.<sup>10</sup> Any statement concerning the nature of one region would be applicable to any other. Unfortunately, in the nature of market networks even in the conceptual construct, such regions do not exist. Given this Lösch mapping, it follows that regional boundaries are not clear-cut and any statement concerning

<sup>6</sup> *Ibid.*, p. 257.

<sup>6</sup> See Lawrence Klein, *Econometrics* (Chicago: Row, Peterson & Co., 1953).

<sup>7</sup> North's consideration of the possible outlets of a region's indigenous savings suggests these considerations (*op. cit.*, p. 255).

<sup>9</sup> August Lösch, *The Economics of Location* (New Haven: Yale University Press, 1954).

<sup>10</sup> In terms of set theory, this implies that the over-all area can be partitioned into disjoint subsets which map one-to-one into each other.

the importance of exports must keep this in mind.

In view of our inability to construct an "ideal" region, the selection of regional boundaries rests on other criteria. Usually, the regional boundaries are suggested by the variables one chooses to study. Non-economic considerations, such as the availability of data and the location of political divisions, may, of course, be the basis for the demarcation of a region. The important point is *not* which boundaries are chosen but the effects of this choice on the variables under study. If the researcher is aware at least of the direction of changes in the variables as a function of regional boundaries, the question of boundaries is of less importance. For example, increased regional size, with more internal trade, implies that the quantitative importance of exports decreases.

Perhaps the most surprising feature of the concept of the export base of a region is that no one, to my knowledge, has attempted to integrate this concept into the traditional foreign-trade multiplier analysis. The works of Metzler, Machlup, and Stolper are conspicuous by their omission from the discussions.<sup>11</sup> Usually the economic base of a region of any size from an urban area up to several states is merely assumed to be exports. Implicitly, no foreign-trade multiplier feedback is assumed. This is probably valid for smaller areas, but for larger areas the feedback can be an important factor. An example may illustrate this point.

Consider the exports of New England. Like those of any other region, its exports compete with products from elsewhere. Thus one expects and finds that export receipts fall off as a function of distance.<sup>12</sup> Few of the region's exports enjoy a worldwide market. The New York area would be

expected to absorb a much higher percentage of New England's exports than would a market of equal size in the Far West. Conversely, the New England area would tend to absorb a greater percentage of New York's exports than would a more remote market.

Contrast this situation with that of a mill town. Here the exports may be considered as going off into some distant space. The income of the mill town will be affected by the income of its market, but the income of the market will not be affected by the income of the mill town. This merely places the mill town in the same position as the competitive wheat farmer who is too small to affect the market but is affected by it. In this case there is no foreign-trade multiplier feedback. This is not true in the regional case, and one is left in the uncomfortable position of having exports in part a function of domestic income. Thus in the short run it appears that the determination of regional income depends only in part on the region's exports. The larger the region under consideration, the smaller the role of exports. Other variables in the structural equations must be considered if income stability is to be more fully understood.

### III

The concept of the export base, or even the fuller concept of regional income determination which includes other autonomous variables, is a short-run concept. As such it may be fairly accurate. Our knowledge of consumer behavior and the relative ease of entry into residentiary activities, such as baking and retailing, indicate that this may be a fairly safe assumption, at least for small regions. To extend this relationship to the question of regional development, however, can be dangerous.

Before we consider the question of the export base in regional development, one issue should be cleared up. It involves a dif-

<sup>11</sup> Lloyd A. Metzler, "Underemployment Equilibrium in International Trade," *Econometrica*, X (April, 1942), 97-112; Fritz Machlup, *International Trade and the National Income Multiplier* (Philadelphia: Blakiston Co., 1943); Wolfgang Stolper, "The Volume of Foreign Trade and the Level of Income," *Quarterly Journal of Economics*, LXI (February, 1947), 285-310.

<sup>12</sup> See Walter Isard and Merton Peck, "Location Theory and International and Interregional Trade Theory," *Quarterly Journal of Economics* LXVII (February, 1954), 97-114.

rence between regional growth and economic development in general. Suppose that we assume that general economic development means raising the per capita income of the area, say North America. Further, let us define regional growth as the rate of change of per capita income in some segment of this totality, say Canada. It is pertinent to ask whether these should be considered as presenting the same sort of problem. If we imagine that the continent had developed without Western influence, it assuming capitalism, some process of primary, secondary, and tertiary evolution might be expected to have taken place just as it did, in general, in the development of Europe. True, some areas might have specialized in agricultural activities; but if the concept of regional balance means anything, specialized areas of manufactures would be expected. If some island economy, unknown to the rest of the world, were studied as a case of economic development, the stage concept might be quite valid.

This sort of analysis should not be called on, however, to deal with questions of regional economic growth, which presents a different sort of problem. If a new peninsula were formed off the New Jersey coast, it could provide an ideal setting for studying regional economic growth. In this case there is no reason to expect the peninsula area to pass from the primary-subsistence to the secondary-tertiary stage as real incomes increase. If, as North points out, the region can develop an export base, it may develop a variety of forms. It could become a center for truck gardens (primary); a site for manufacturing (secondary); or a vacation area (tertiary). Note that this does not imply that it will develop even if it *seems* to have an export base, for reasons to be discussed later. The important point about the New Jersey example is that we are dealing with a region in the neighborhood of more advanced areas. The degree of specialization and of exporting will depend on the market. The higher the incomes in the neighboring areas, given the propensity to import, the higher the volume of their imports, that is,

the exports of the peninsula. The volume of exports and, in turn, internal growth will depend not only on the factor endowment but also on the income of the surrounding area. It is useful to keep this distinction in mind when contrasting regional economic growth with economic development in general.<sup>12</sup>

The idea that essentially the export base is the necessary and sufficient condition for regional economic growth may be, by definition, a true statement. Given the transport network, the size and location of markets, and factor endowments, it appears that a region will develop if it can compete with other regions in the export market. This implies an ability to produce at lower cost. With factor mobility, growth will take place only if the return to the factors is equal to, or greater than, the return to the same factors in other regions. If this is what is meant by the ability to develop an export base, it is correct by definition, but it does not uncover enough to predict growth. Ability to find an export base depends not only on the value of the units of output but on the cost of the inputs. These costs cannot be assumed to be equal for all regions. Yet, if residentiary activities are assumed to be endogenous and are not considered as a factor in regional growth, the analysis will implicitly assume that all unit factor costs are equal among regions. Put another way, it is possible to define the necessary condition for regional economic growth as the creation of an export base. But location theory, which is called on to explain its creation, will work only if factor costs are known. The determination of factor costs depends in part on the nature of the region's residentiary activities.

<sup>12</sup> The development of the Pacific Northwest and the Canadian development cited by North (*op. cit.*, p. 246-47) may be analogous to the New Jersey example. Both occurred after the process of industrialization was under way. In contrast to this case, the position of the earliest colonies typifies the case of general economic development. Of course, this is a matter of degree and should not be taken as a statement that exports were unimportant to the early colonies.

An example may serve to illustrate this point. Going back to our hypothetical New Jersey peninsula, assume that a coal deposit is found some two hundred miles out on the peninsula. Will it be mined to compete in the New York market with Pennsylvania coal? Make one further assumption about the region. Assume that the rest of the area is all sand and marshland. If workers are to mine this newly found deposit, they must eat, and hence there must be imports. If the cost of these imports is high enough, no coal will be mined, and no export base will develop.

Contrast this with a situation in which the peninsula is rolling, fertile countryside. Truck gardening and dairy farming can develop. Some imports will still flow in, but some local needs—vegetables and milk—will be supplied locally, that is, supplied by residentiary activities. Under these conditions coal may be mined because of the lower cost of production, in this example lower dollar wages.<sup>14</sup>

Again, formally speaking, it is the ability to develop an export base which determines regional growth. Yet in terms of causation, the nature of the residentiary industries will be a key factor in any possible development. Without the ability to develop residentiary activities, the cost of development of export activities will be prohibitive.

The objection may be raised that this is a special case. No claim is made for general validity. However, if one seeks to explain the failure of certain parts of Alaska or Canada to develop, this consideration may uncover a more complete picture. Further, it is well known that cities usually develop in locations that are surrounded by good lands and not in the middle of less fertile areas.

<sup>14</sup> In both cases the real wages of the coal miners would be the same, but in the former case, because of the high cost of living, dollar wages would be higher. In location theory it is the dollar cost which determines location and development.

However, the idea of the export base is more useful when applied to certain areas, such as satellite cities in the suburban fringe. Here low transport costs and proximity to markets insure that, even if residentiary activities do not develop fully, their outputs can be imported from near-by areas. The larger the region under consideration, the less safe the assumption.

A final point is in order concerning regional growth and the ratio of export to residentiary activity. Given its population, boundaries, transport network and costs, markets, and factor endowment, a region must divide its energies between residentiary and export activities. If too little is devoted to one or the other, the economy will not be maximizing per capita income. Supposedly there is some optimum division. If export activities are relatively too large, it will pay to move resources into residentiary industries (witness the enviable position of the storekeeper during the gold rush), and the region's income will increase. Here we find an example in which regional growth is possible with a reduction of exports.

#### IV

This note has tried to show that the concept of the export base is merely one aspect of a general theory of short-run regional income determination. In the case of large regions, other variables may play as important a role as exports. Furthermore, the concept of the export base may be useful in describing regional income growth, but this need not be considered the same problem as general economic development. As an explanatory factor in regional growth, the idea of the export base should not subsume the key role of residentiary activities in determining factor costs of possible regional exports. Finally, since a region must optimize the use of factors as between exports and residentiary outputs, a decline in export activity may even be accompanied by rising regional income.

## A REPLY

DOUGLASS C. NORTH<sup>1</sup>

PROFESSOR TIEBOUT's comment is a welcome addition to the literature on regional economic growth. The role of the export base in regional development requires further analysis, and Tiebout has raised some important questions that merit discussion.

The bulk of his criticism of the role of the export base in regional growth hinges on one critical point at issue between us. His is a *short-run* analysis, in which the export base is conceived to be only one of a number of important factors in income determination. I have no quarrel with this position, but it has little relevance for my article, which was explicitly concerned with *long-run* economic growth. Short-period income determination and long-run economic growth are not the same thing. In the former case the analysis is concerned with changes in the level of employment and the variables that will affect the rate of utilization of productive factors. In this case increased business investment will result in expanded employment and income in periods of less than full employment. Such analysis, however, has little relevance for long-run economic growth, where the objective is to determine the factors that will affect the decade-to-decade changes in the real aggregate and per capita income of an area under conditions of full employment. In the latter case secular expansion comes about because of increased output per unit of resources or an increase in the supply of productive factors, or both. Historically, this increase in labor and capital has come about as a result of long-run expansion of the demand for productive factors within

the area. Not only has there been mobility of productive factors within the American economy, but also during a substantial period of our growth there has been international mobility of capital and labor for the entire Atlantic economy. Therefore, while the study of short-run income determination has been concerned with the rate of utilization of productive factors, the study of long-run growth has dealt with the determinants of changing efficiency and the immigration of labor and capital into an area. The variables used in income analysis are of limited use in the study of long-run growth. Indeed, the aggregates used tend to obscure rather than to illumine the factors generating secular expansion. An examination of Tiebout's major points will further clarify this distinction and highlight some important problems for further research.

## I

Tiebout and I are in agreement that there is no "ideal" region. Since he concedes that the question of boundaries is of less importance if the researcher is aware of the significance of increased regional size, there is no apparent difference in our position. Yet one point requires emphasis. The usefulness of a region as an economic unit of analysis rests upon its specialization. It is this geographic division of labor, with different areas having special factor endowments and transfer costs, which makes the concept of the region valuable in economic analysis. The region's significance lies in its being a specialized part of the whole. If the size of the region is to be limited by its individual economic characteristics, then the concept of a geographically contiguous area held together by its development around a common export base is a useful (though certainly not the only) basis of classification.

<sup>1</sup> I am indebted to my colleagues Philip Cartwright and Donald Gordon for suggestions which have clarified some of the points at issue in this discussion.



It has the added advantage that, in terms of the long-run growth of different areas in America, the export base has been influential in shaping a good deal of the history of the region.<sup>2</sup> Such a classification necessarily limits the size of the region and minimizes the problems raised by Tiebout.

## II

Given the region as defined, the role of the export base in regional growth may now be more precisely delineated. Tiebout and I are in agreement that it is not the sole source of regional growth,<sup>3</sup> but we are in disagreement when he states that it may not even be the most important factor in regional expansion. An examination of the differential rates of growth of regions throughout America's development inevitably focuses on the ability of areas to attract productive factors. Initially it was the rich land and resources capable of producing extractive goods in demand in existing markets which were the primary attractions. At a later date, with changing factor combinations and technological developments, it was frequently the opportunities in manufacturing for the United States market which led to immigration of labor and capital into a region. The important point is that the pull of economic opportunity as a result of a comparative advantage in producing goods and services in demand in existing markets was the principal factor in the differential rates of growth of regions.

Since residentiary industry depends on income within the region, the expansion of such activity must have been induced by the increased income of the region's inhabitants. Therefore, increased investment in residentiary activity is primarily induced investment as a result of expanded

income received from outside the region, and, correspondingly, expanded employment in locally oriented industry, trade, and services primarily reflects long-run changes in income received from the export base.<sup>4</sup> The qualifications to this argument require examination in order that the significance of the export base in regional growth may be properly evaluated.

1. Disproportionate federal government expenditures in a region (as compared with tax withdrawals) can serve and have served as a cause of regional expansion. Not only may the character and amount of federal expenditure in a region be expansive, but also investment in social overhead facilities in a new area may alter its competitive position with other regions.

2. Migration for non-economic reasons may lead to expansion of residentiary activity without any expansion of the export base. In a high-income society such as ours the lure of pleasant living conditions—"amenities," to use Professor Ullman's term<sup>5</sup>—has been a force attracting immigrants (with capital) into California and other areas and leading to an expansion of residentiary activities.

3. The relationship between residentiary activity and imports changes in the course of regional development. With the opening-up of a new area, almost everything must initially be imported. Gradually, residentiary activity increases until locational factors

<sup>4</sup> The employment multiplier has been conceived by Hildebrand and Mace ("The Employment Multiplier in an Expanding Industrial Market: Los Angeles County, 1940-47," *Review of Economics and Statistics*, Vol. XXXII, No. 3 [August, 1950]) to be primarily of use in short-run analysis. However, the study by the Federal Reserve Bank of Kansas City of "The Employment Multiplier in Wichita" (*Monthly Review, Tenth Federal Reserve District*, Vol. XXXVII, No. 9 [September 30, 1952]) strongly suggests that residentiary employment does not adjust to short-run changes in employment in export industries but does reflect long-run movements in export employment.

<sup>5</sup> Edward Ullman, "A New Force in Regional Growth," *Proceedings of the Western Area Development Conference, November 17, 1954* (Palo Alto, Calif.: Stanford Research Institute, 1955).

<sup>2</sup> A brief account of the role of the export base in shaping the character of a region's economy is given in my article, "Location Theory and Regional Economic Growth," *Journal of Political Economy*, LXIII (June, 1955), 249-51.

<sup>3</sup> See the qualifications in my article, *ibid.*, p. 250, n. 34.

effect a balance between imports and locally oriented economic activity at a given level of technology and transfer costs. Since techniques and transportation have undergone radical changes over time, this relationship has been subject to important changes. Moreover, as a region's population and income grow, its regional market will become large enough to make it feasible to produce some goods and services locally which had previously been imported.<sup>6</sup>

Clearly, therefore, residuary activity does not play a purely passive role in regional growth. Tiebout's point about factor costs is a good one. Both the nature of the supply curve of labor and the level of transfer costs are important determinants of the ability of a region to produce export commodities. Typically, new regions have been opened up and developed because they had such a tremendous advantage in natural-resource endowments that they could produce and market their export commodities at a cost competitive with other areas despite this disadvantage in labor and transfer costs (and in the case of the earlier development of American regions, when institutions for financial mediation were immature and capital was less mobile, higher capital costs as well). The subsequent inflow of capital and labor and the development of social overhead facilities typically reduced these cost disadvantages and made it possible to produce other export commodities whose comparative resource advantages were somewhat less pronounced. However, when the growth of residuary industry is "stunted" and transfer costs remain high, then the export base will not expand in this manner. Alaska is an excellent case in point.

This examination of the major alternative sources of regional growth clearly indicates their secondary importance as compared with the export base in long-run regional growth. The first two qualifications are

<sup>6</sup> The changing character of residuary activity with regional growth requires further research both in expanding regional markets and in the historical development of regions where changing technology and transfer costs have changed the character of residuary goods and services.

exceptional in character, while the third, though more important, is clearly not a primary determinant of growth but rather a factor that will exert some influence upon a region's rate of growth.<sup>7</sup> Any analysis of the secular growth of a region must be primarily focused on the success of its export base, and Tiebout's contribution here has been to point out some of the factors that must be considered.<sup>8</sup> However, his discussion of short-run factors in income determination, significant though these may be in determining the level of employment (or effecting a shift in resources in the case of full employment), is irrelevant to the analysis of long-run regional economic growth.

### III

Tiebout's distinction between regional growth and economic development in general is surely a spurious one. It is hard to conceive of the economic growth of one region which would not favorably affect the per capita income of the nation as a whole (even though it might have adverse effects upon another region). In fact, America's entire development has illustrated this relationship. Growth has been generated in particular geographic areas which, as a result of favorable factor endowments and transfer costs, could produce goods in demand in existing markets. Whether it was the opening-up and development of a new region in the West capable of producing wheat for the world market or the development of an industrial region in the eastern and central states producing manufactured goods for the domestic market, the result in each case was to attract labor and capital (from Europe as well as internally) not only

<sup>7</sup> Tiebout's final point deals with the possibility of regional growth with a decline in exports. This is conceivable but flies in the face of the experience of growing nations, where international trade has typically increased with rising incomes rather than the reverse.

<sup>8</sup> In this regard the changing terms of trade of a region have been important. Regions whose export base consists primarily of agricultural commodities have been particularly affected.



into the expanding export industry but also into a wide variety of residentiary activities to meet the expanding needs of the region's population. The process of urbanization, which was an integral part of the growth of manufacturing regions, was as expansive in its effects as was the opening-up of new regions, their development, and their assimilation into the economy. In both cases the expansion of the region required a vast increase in imports from outside the area. The result was to induce investment throughout the rest of the economy. The multiplier-acceleration process that resulted was an essential part of America's economic growth.<sup>9</sup>

Tiebout's footnote about the difference between the later development of Canada

<sup>9</sup> James S. Dusenberry has an excellent account of this entire process in his article, "Some Aspects of the Theory of Economic Development," *Explorations in Entrepreneurial History*, Vol. III, No. 2 (December, 1950).

and the Pacific Northwest, when there were well-developed markets, and the case of general economic development which typified the American colonies is, to the best of my knowledge, likewise incorrect. America was settled partly for the explicit purpose of producing goods in demand in the expanding European market. The prosperity of the colonies did not rest upon subsistence farms but resulted from the rich land and resources of the New World, which could produce tobacco, rice, indigo, ships, fish, cereals, and other products that were in growing demand in England, Europe, and the West Indies. The whole development of the New World has been within the context of the rapidly expanding Western world, and the prosperity of the colonists reflected the growth of income throughout the North Atlantic economy, which resulted in an expanding demand for their services (particularly important in the case of New England) and commodities.

## REJOINDER

CHARLES M. TIEBOUT<sup>1</sup>

PERHAPS the basic issues between Professor North and myself can be easily resolved. It is my contention that the concept of the "export base" is merely an oversimplified method of describing regions by structural equations of the national-income or general-equilibrium variety. The larger the region, the greater the oversimplification in designating exports as the exogenous variable. This is clearly true in the short run.

For long-run growth, merely to look at exports as the key factor in explaining regional growth is no more adequate than merely looking at investment at the national level. Yet if econometric models of national trends are useful in growth analysis, there is no a priori reason why the same type of more complete system cannot be applied to regional growth.<sup>2</sup> North is concerned about the factors determining the level of export activity. My concern, however, is not only with the factors affecting the level of export activity but with the level of other exogenous variables as well. Further, given their level, what does the nature of the structural relations between the exogenous and the endogenous or residentiary activities imply for total regional activity?<sup>3</sup>

Whether or not my classification of regional growth as opposed to general econom-

ic development is "spurious" is a matter of both judgment and degree. Unfortunately, no pure cases as such have ever occurred historically. Yet a fundamental difference does seem to appear. In describing the development of all the region known as western Europe, one is dealing with a substantially *closed economy*.<sup>4</sup> Development proceeded by stages, since there was no alternative. In the nature of the hierarchy of wants, it would have been impossible to pass directly to the tertiary stage.

The regional economy, which is an *open economy*, need not pass through the stage process. The region Florida, for example, may be considered as having started largely in the tertiary stage. The ability of a region to start at any stage, given factor endowment, depends on two conditions: (1) the regional economy must be open; (2) the region must deal with other regions which are able to supply it with products of other stages that it does not produce.

Thus one should not expect any particular region to pass through the three stages of development. However, the reason is not that the stage theory is necessarily wrong but simply that in some cases it may not apply.

<sup>2</sup> North's contention that "the aggregates used tend to obscure rather than to illumine the factors generating secular expansion" (p. 165) seems to overlook this aspect of growth analysis.

<sup>4</sup> What is a closed or an open economy is also a matter of degree and may be considered in terms of the proportion of exports to total income.

<sup>1</sup> I am grateful to my colleague Bernard Kemp for sharpening some of the points involved.

<sup>2</sup> See Stefan Valavanis-Vail, "An Econometric Model of Growth: U.S., 1869-1953," *American Economic Review*, XLV (May, 1955), 208-21.

# 15 Industrialization, Factor Markets, and Agricultural Development<sup>1</sup>

William H. Nicholls<sup>2</sup>

Reverse cannot befall that fine Prosperity  
Whose sources are interior.

EMILY DICKINSON

## I. INTRODUCTION

FOR nearly a decade, we have had under way at Vanderbilt University a large-scale research project on Southern Economic Development and Agriculture. This project has been concerned with testing for selected southern areas the locational hypothesis first advanced by T. W. Schultz in 1950 to explain wide international and interregional divergences in per capita incomes.<sup>3</sup> As Schultz himself recognized, it would be virtually impossible to test his hypothesis on an international basis because of

the vast number of uncontrollable variables.<sup>4</sup> Instead, Tang and I have dealt with relatively small and homogeneous areas for which one might hope to isolate more successfully the effects of industrial-urban development (via factor

<sup>1</sup> A revision of a paper prepared for the Conference on the Role of Agriculture in Economic Growth, sponsored by the Social Science Research Council's Committee on Economic Growth, at Stanford University, November 11-12, 1960.

<sup>2</sup> I am greatly indebted to my colleague, Anthony M. Tang, on whose own contributions to the present subject I have drawn so freely in what follows as to give him the status of virtually a joint author. I am also grateful to the Rockefeller and Ford Foundations for the financial support which made our various studies possible.

<sup>3</sup> T. W. Schultz, "Reflections on Poverty within Agriculture," *Journal of Political Economy*, LVIII (1950), 1 ff.; "A Framework for Land Economics—the Long View," *Journal of Farm Economics*, XXXIII (1951), 204-15; and *Economic Organization of Agriculture* (New York: McGraw-Hill Book Co., 1953), chaps. ix and x. According to Schultz's hypothesis, economic development occurs in a specific locational matrix which is primarily industrial-urban in composition and near the center of which the existing economic organization (factor and product markets) works best. He further hypothesized, with particular reference to agriculture, that those parts of agriculture situated favorably relative to the industrial-urban center enjoy more efficient factor and product markets than those situated at the periphery.

<sup>4</sup> However, as Meier and Baldwin (*Economic Development: Theory, History, Policy* [New York: John Wiley & Sons, 1957], p. 147 and Part II) have shown, the mid-nineteenth-century world economy can be rather successfully analyzed in the general terms of center (Britain) and periphery.

Reprinted from the *Journal of Political Economy*, vol. 69, no. 4 (August 1961), pp. 319-340, with permission of the author and The University of Chicago Press. © 1961 by The University of Chicago.

markets) on agricultural productivity and income. Tang has presented most of his findings in his path-breaking book on industrial-urban development and agriculture in the South Carolina-Georgia Piedmont.<sup>5</sup> The findings of my parallel study of the Upper East Tennessee Valley have been reported in a long series of journal articles.<sup>6</sup> Since our investigations are now virtually complete, the purpose of the present paper is to bring together within the compass of a single article the principal generalizations which have emerged from our joint efforts.

*Objectives and methodology.*—While our approaches have differed in detail, Tang and I have each sought to test a series of subhypotheses which may be grouped into three major hypotheses: (1) that the labor, capital, and product markets facing agriculture are relatively more efficient in local areas which have enjoyed considerable industrial-urban development than in similar nearby areas which

have not; (2) that the effects of differences in rates of industrial-urban development have accordingly been increasingly wide inter-area differences in per worker agricultural capital and farm output (income); and (3) that there are fundamental impediments to equilibrating factor movements which tend to perpetuate differences in market efficiency and agricultural productivity between the more and less industrial areas. To this end, each of us selected a group of twenty to twenty-one contiguous counties (each with a combined area equivalent to that of Massachusetts or Connecticut) which have had a common historical and cultural background; which some fifty to seventy-five years ago had a similar dependence on agriculture and similar farm output per worker; but some of which have since experienced substantial rates of industrialization while others (the control group) have remained largely rural-agricultural.

Each of us has made a very thorough analysis of the economic development of his area during the entire century 1850–1950, but we have put greater emphasis upon the period since 1900 and have analyzed with special intensity the dynamic years since 1940. For sources of data, we have relied primarily upon the United States Census, supplemented insofar as possible from other important statistical compendiums which contained relevant socioeconomic data on a county-unit basis. Our statistical methodology has been relatively simple, consisting largely of correlation techniques applied to the ranks of our data rather than to the original data themselves. Tang successfully employed Wilcoxon's rapid sum-of-ranks technique which often enabled him to eliminate calculations of various socioeconomic indexes for all but his five top-ranking (most industrial) and five bot-

<sup>5</sup> Anthony M. Tang, *Economic Development in the Southern Piedmont, 1860–1950: Its Impact on Agriculture* (Chapel Hill: University of North Carolina Press, 1958).

<sup>6</sup> In order of their publication, my own journal articles were: "A Research Project on Southern Economic Development, with Particular Reference to Agriculture," *Economic Development and Cultural Change*, II (1952), 190–95 (which set forth in detail the various hypotheses of the entire project); "Some Foundations of Economic Development in the Upper East Tennessee Valley, 1850–1900," *Journal of Political Economy*, LXIV (1956), 277–302, 400–415; "The Effects of Industrial Development on Tennessee Valley Agriculture, 1900–1950," *Journal of Farm Economics*, XXXVIII (1956), 1636–49; "Human Resources and Industrial Development in the Upper East Tennessee Valley, 1900–1950," *Quarterly Journal of Economics*, LXXI (1957), 289–316; "Relative Economic Development of the Upper East Tennessee Valley, 1850–1950," *Economic Development and Cultural Change*, V (1957), 308–24; "Industrial-Urban Development and Agricultural Adjustments, Tennessee Valley and Piedmont, 1939–54," *Journal of Political Economy*, LXVIII (1960), 135–49; and "Factors Affecting Gross Farm Income Per Worker in the Upper East Tennessee Valley, 1899–1954," *Journal of Farm Economics*, XLII (1960), 356–62.

tom-ranking (least industrial) counties. Because of the more erratic nature of several of the counties in my study area, I found the Wilcoxon method less satisfactory. Instead, I applied Spearman rank-correlation techniques to my entire sample of twenty counties, in each case correlating their ranks in industrial development<sup>7</sup> with their ranks in numerous other socioeconomic indexes. While our choice of these simpler statistical methods was primarily based on their economies in computing time, their use also appeared to be appropriate in view of the crudeness of our data, the small numbers of our observations, and the failure of our data to fulfil the assumptions implicit in parametric techniques.

In the remainder of this paper, I shall present a summary of the generalizations which emerge from these two case studies. However, because Tang's statistical methods differed substantially from mine, I shall normally not present the statistical findings of the two studies comparatively. Instead, I shall state various generalizations in strictly verbal terms, using only my own statistical results for support or illustration. In each case, I shall also indicate in general terms whether Tang's findings are in agreement or in conflict with my findings. I shall begin with a section on the pre-industrial period, when any differences in agricultural productivity and incomes might be

expected to have reflected differences in "original" physical resources. I shall then turn to the more recent period since 1900, during which both study areas were subject to the dynamic effects of substantial industrial-urban development. For the latter period, I shall center attention on the responses to industrialization of capital and labor markets serving the nearby agriculture. First, however, let us gain some perspective by looking briefly at the major patterns which emerged during the ninety years 1860-1950.

## II. SOME COMMON PATTERNS IN INDUSTRIAL-URBAN DEVELOPMENT, 1860-1950

*The pre-industrial period.*—Despite many significant differences in their natural and socioeconomic attributes, the agricultures of the Tennessee Valley and Piedmont study areas shared one important characteristic—the tendency for intra-area geographical differences in per worker farm income and farm capital to be substantially reduced during 1860-1900 and to be markedly increased during the half-century beginning in 1900. This common agricultural pattern of initial convergence and subsequent divergence corresponded roughly with an initial, essentially pre-industrial period, followed by a period of rapid and notable industrial-urban development within each area. These long-run tendencies may be illustrated by some summary indexes taken from my Tennessee Valley study and presented in Table 1.

During 1860-1900, per capita value added by manufacture in the Tennessee Valley never exceeded its peak of 17 per cent of the national average attained in 1860. In fact, relative to the nation as a whole, the Tennessee Valley underwent significant industrial retrogression during 1860-90 (reaching a low of 8 per cent in 1890) and had not quite recovered its

<sup>7</sup> In our earlier work, Tang and I used "per capita value added by manufacture" as our measure of industrial development since the necessary basic data were reported by the Census from 1860 on. However, beginning in 1930, the Census reported data permitting the computation of a moderately comprehensive index of "per capita non-agricultural payrolls" (manufacturing, retail and wholesale trade, and selected service industries), which was almost invariably more highly correlated with other socioeconomic indexes than was per capita value added. Hence, here as in our other more recent work, I have adopted per capita payrolls as my measure of industrial-urban development.

# INDUSTRIALIZATION AND AGRICULTURAL DEVELOPMENT 361

1860 level by 1900. The Piedmont area did not show the same relative industrial retrogression and, during 1880-1900, gained sharply on the rest of the nation's manufacturing with the expansion of its new textile industry. Nonetheless, even the Piedmont's per capita value added

have greater per capita values added than did today's less developed counties. (In fact, the corresponding coefficient of rank correlation, .432\*, for 1860 indicates that even forty years earlier this particular index had some predictive value for the relative development of the

TABLE 1

SOME SUMMARY INDEXES OF INDUSTRIAL-URBAN AND AGRICULTURAL DEVELOPMENT, UPPER EAST TENNESSEE VALLEY STUDY AREA, SELECTED YEARS, 1860-1954\*

|  | 1860  | 1880   | 1900   | 1920   | 1940   | 1950   | 1954   |
|--|-------|--------|--------|--------|--------|--------|--------|
| Per capita value added by manufacturing: |       |        |        |        |        |        |        |
| Index for area (U.S. average = 100)..... | 17%   | 11%    | 15%    | 21%    | 55%    | 62%    | 68%    |
| Coefficient of variation.....            | 0.87  | 0.97   | 0.95   | 0.94   | 1.63   | 1.43   | N.a.   |
| Coefficient of rank correlation.....     | .432* | .650*  | .878** | .868** | .869** | .911** | .979** |
| Population per square mile:              |       |        |        |        |        |        |        |
| Index for area (U.S. average = 100)..... | 242%  | 213%   | 187%   | 157%   | 162%   | 155%   | N.a.   |
| Coefficient of variation.....            | 0.24  | 0.26   | 0.22   | 0.29   | 0.47   | 0.62   | N.a.   |
| Coefficient of rank correlation.....     | .236  | .114   | .280   | .644** | .774** | .792** | N.a.   |
| Per worker gross farm income:            |       |        |        |        |        |        |        |
| Index for area (U.S. average = 100)..... | N.a.  | 58%    | 58%    | 52%    | 53%    | 45%    | 39%    |
| Coefficient of rank correlation.....     | N.a.  | .268   | .185   | .186   | .678** | .624** | .744** |
| Per worker farm capital:                 |       |        |        |        |        |        |        |
| Index for area (U.S. average = 100)..... | 95%   | 66%    | 49%    | 43%    | 56%    | 59%    | 50%    |
| Coefficient of rank correlation.....     | .488* | .568** | .344   | .484*  | .808** | .798** | .841** |

\* Coefficient of correlation between county ranks in the particular index in each selected year and the county ranks in my index of economic development, per capita non-agricultural payrolls in 1954.

\* Significant to 5 per cent level.

\*\* Significant to 1 per cent level.

stood at only 30 per cent of the national average in 1900.<sup>8</sup> To be sure, there were in 1900 already considerable *relative* differences in the levels of industrialization of the several counties within each area, as illustrated by the fact that there was a very strong tendency (.878\*\*) for today's more industrial-urban counties to

several counties a century later.)<sup>10</sup> Despite these significant relative differ-

<sup>9</sup> Here and in subsequent analysis, the figures in parentheses will (unless otherwise specified) represent the coefficients of rank correlation between my index of industrial-urban development (per capita non-agricultural payrolls) for 1954 and the particular socioeconomic index and year indicated in the text, for the twenty counties of the Tennessee Valley study area. (One asterisk indicates statistical significance at the 5 per cent level; two asterisks, significance at the 1 per cent level.)

<sup>8</sup> While both areas had per capita values added which were 11 per cent of the national average in 1880, the Piedmont's industrial progress during 1880-1900 was obviously much greater—standing at 30 per cent as compared with the Tennessee Valley's 15 per cent in the latter year.

<sup>10</sup> That the same was true for the Piedmont area is emphasized by the indexes of Table 8 in Tang, *op. cit.*, p. 88; see also pp. 60-63.

ences, it seems clear that as late as 1900 (a) all of the counties within each area were primarily agricultural (the estimated proportions of the male work force engaged in agriculture in the twenty Tennessee Valley counties ranged from 69 to 92 per cent) and (b) the absolute differences in the degree of their industrialization were as yet not sufficiently great to cause substantial differences in local agricultural productivity and incomes.

Furthermore, insofar as industrial development had taken place during 1860-1900, its unevenness among the several counties of each area was not yet marked (the coefficient of variation in per capita values added during 1860-1900 increased only from .87 to .94). In addition, inter-county differences in population density had, if anything, slightly declined (the corresponding coefficients of variation were .24 and .22), and as late as 1900, for the Tennessee Valley at least,<sup>11</sup> today's more advanced counties did not have significantly greater population densities (the coefficient of rank correlation was only .280) than did today's less developed counties. Hence, our choice of 1900 as the approximate date ending the pre-industrial period of each area would appear to be reasonable enough.

*The period of industrial development.*—In any case, during the half-century following 1900, both the Tennessee Valley and the Piedmont experienced rates of industrial development far in excess of the nation as a whole. Between 1899 and 1947, the per capita value added by manufacture in the Tennessee Valley area increased from 15 to 62 per cent of the national average (Table 1); in the

Piedmont area it increased from 30 to 120 per cent.<sup>12</sup> Within each area, however, the various counties shared far from equally in this industrial growth. For example, for the Tennessee Valley area, the coefficient of variation in county per capita values added increased from 0.95 to 1.63 during 1899-1939, although it then declined to 1.43 in 1947. The corresponding coefficient of variation in county population densities increased steadily from 0.22 to 0.62 during 1900-1950. Furthermore, although today's more industrial counties did not have significantly greater population densities (.280) in 1900, by 1950 they showed a very strong tendency to do so (.792\*\*). Whereas, in 1900 only one of the Tennessee Valley's twenty counties (and five of the Piedmont area's twenty-one) had per capita values added by manufacture exceeding the modest level of 40 per cent of the national average, by 1947 seven (and fifteen) counties did so. In fact, by the latter year, two (and nine) counties actually exceeded the national average. Thus, industrialization (and concomitant urbanization) proceeded at a rapid but highly uneven pace within each study area.

*Trends in agricultural income differentials.*—Finally, as Table 1 makes clear for the Tennessee Valley, within each study area there was a convergence in county indexes of per worker farm capital and income during 1860-1900, fol-

<sup>11</sup> In the Piedmont area, today's more industrial counties already had significantly greater population densities (although absolute differences were still small) in both 1850 and 1900 (Tang, *op. cit.*, p. 97).

<sup>12</sup> While the Piedmont's industrialization far outstripped that of the Tennessee Valley, the latter had the advantage of greater diversification of industries and greater relative male employment in manufacturing. As late as 1950, textiles accounted for 76 per cent of the Piedmont's manufacturing employment; the Tennessee Valley's leading industry, chemicals, accounted for only 29 per cent, followed by textiles and apparel (22 per cent) and furniture, lumber, and wood products (19 per cent). In the same year, females were responsible for 40 and 24 per cent, respectively, of total manufacturing employment in the two areas.



lowed by an increasingly divergent movement during 1900-1950. (Here we are concerned primarily with intercounty differences, although the generally unfavorable trend in farm capital per worker relative to the national average may be noted.) Thus, prior to 1900, today's more industrial counties showed a tendency to have higher per worker farm capital and income than their less industrial neighboring counties, but this tendency was weakest in 1900. During the subsequent period of substantial industrial-urban development, however, this tendency had become much stronger than at any other time during the entire century.<sup>13</sup> To what extent can we attribute the increasingly favorable position of the more industrial counties after 1900 to their more rapid rates of industrialization? Before seeking to answer this question, let us consider the pre-industrial period 1860-1900 and particularly the extent to which today's more advanced counties owe their superior achievements both agricultural and industrial to superior "original" resource endowments.

### III. FACTOR ADJUSTMENTS IN THE PRE-INDUSTRIAL YEARS, 1850-1900

*Differences in "original" resource endowments.*—Within both the Tennessee Valley and the Piedmont today's more industrial counties had in 1860 an agriculture with certain significant advantages over the agriculture of today's less industrial counties. Among these advantages were superior "original" agricultural land, certain differential windfall gains from new cash crops, and higher rates of capital formation in agriculture. While in 1860 today's more industrial counties did not have any more total acreage of land

per farm ( $-.141$ ), they showed in both areas some tendency to have more improved acreage per farm (.393) and a highly significant tendency to have a higher value per acre of farm land (.639\*\*). In each area, there was some tendency for today's more industrial counties to have in 1860 greater production per thousand acres of improved farm land of some major cash crop—cotton in the Piedmont, wheat (.816\*\*) in the Tennessee Valley. This combination of superior "original" agricultural resources and greater suitability for a profitable new cash crop was capable of giving rise to important intercounty differences in capital formation within agriculture. Thus, in both areas in 1860 today's more industrial counties had significantly more farm capital per worker (.488\*<sup>14</sup>).

The fact that in each area today's more and less industrial areas showed certain significant differences in 1860 may at first seem surprising. It should be remembered, however, that both areas were approaching the end of their first century of settlement, so that enough time had already passed for differences attributable to differential "original" resource endowments to have become apparent. Initially, all counties had had a self-sufficing agriculture in a still largely unsettled and uncleared wilderness so that "original" differences in land productivity were of less importance. As time went on, however, those people who had settled in counties with less productive land found themselves increasingly disadvantaged. Their continuing self-sufficiency perpetuated their economic and cultural isolation and made it difficult for them to finance out of agricultural sur-

<sup>13</sup> For a discussion of the parallel pattern of convergence and divergence in the Piedmont, see Tang, *op. cit.*, pp. 85-90, 69-76.

<sup>14</sup> For the supporting data for the Piedmont, see Tang, *op. cit.*, pp. 29-30, 87-88. Cf. my article, "Some Foundations of Economic Development . . .," *op. cit.*, esp. Table 8 and pp. 293-302.



pluses the beginnings of local industrial development. Today's more advanced counties, on the other hand, tended to have superior "original" agricultural resources and mineral resources (particularly iron ore). Their agriculture was capable of accommodating the use of large quantities of capital and was productive enough after such capital became available to finance some industrial development by 1860. As a consequence, there was already a significant tendency in both areas for today's more developed counties to have a higher per capita value of commercial manufactures (.550\*) than today's less developed counties.<sup>15</sup>

Had the advantage of today's more developed counties in terms of such indexes as per worker farm capital and gross farm income persisted during subsequent decades, we might settle the matter by simply attributing today's wide farm-income differentials to differences in "original" resource endowments. Instead, however, as we have already observed, there was in both areas a convergent movement in county indexes of per worker farm capital and gross farm income during 1860-1900.

*Equilibrating factor adjustments, 1860-1900.*—During the immediate post belum decades, in which industrial development in both areas fell far short of keeping pace with national industrial development, important equilibrating forces seem to have been at work, making today's more and less developed counties more alike than before. Thus, in both areas, today's more advanced counties had significantly higher farm capital per worker in 1860 (.488\*) but not in 1900 (.344).<sup>16</sup> The convergence of gross farm

income per worker is less clear cut, especially in the Tennessee Valley area,<sup>17</sup> but in 1900 today's more advanced counties did not have per worker farm incomes significantly different (.185) from today's less advanced counties in either area. Such convergent movements might have been brought about (a) under conditions of complete immobility of factors of production, if today's more developed counties had had sufficiently more rapid natural population increase to have leveled down differences in per worker farm capital or income; or (b) under conditions of sufficiently great mobility of the factors of production (particularly labor) to eliminate initial differences.

Can the tendency toward convergence be explained by differentially higher rates of natural increase in the counties with superior "original" resources? The evidence supports a negative answer. In the Tennessee Valley at least, today's more industrial counties already showed in 1850 some tendency to have *lower* fertility rates (−.374) and by 1880 the negative relationship had become statistically significant (−.426\*). In both areas, there was also a slight tendency for today's more industrial counties to have in 1850 *smaller* mean family size (−.229),

<sup>16</sup> For the Tennessee Valley area, after reaching a peak of .591\*\* in 1870, this relationship steadily declined to .344 in 1900, when it was at its lowest level for the entire period 1850-1954.

<sup>17</sup> For the Tennessee Valley area, this relationship rose from .098 to .417\* during 1870-90, then fell to .185 in 1900, its lowest level during 1880-1954. During 1880-1900, the relative dispersion of this index (quartile difference divided by quartile sum) declined from .094 to .088 while, for the Piedmont, the coefficient of variation declined much more sharply from .207 to .094. For the latter area, the coefficient of variation of per worker farm capital dropped from .265 in 1860 to .160 in 1900, but, for the Tennessee Valley, the corresponding coefficient increased from .206 to .297 (see Tang, *op. cit.*, pp. 64, 71, 86, 88; and Nicholls, "Some Foundations of Economic Development . . .," *op. cit.*, Table 9, p. 301).

<sup>15</sup> Nicholls, "Some Foundations of Economic Development . . .," *op. cit.*, pp. 414, 281-86, and Tang, *op. cit.*, pp. 57-60, 88.

although this tendency had by 1900 become stronger in the Tennessee Valley ( $-.394^*$ ) while being virtually eliminated in the Piedmont. In any case, it is clear that the tendency toward convergence in per worker farm capital and income during 1850–1900 cannot be attributed to today's more developed counties having had *greater* rates of natural increase than their less industrial neighbors. Hence, we must turn to the evidence on degree of factor mobility for an explanation. Throughout 1850–1900 in the Piedmont, today's more industrial counties showed some tendency to have lower net outmigration rates than today's less industrial counties. While the same was not true initially in the Tennessee Valley, by 1890–1900 this tendency had become statistically significant ( $-.453^*$ ). As a consequence of the joint effects of differential rates of human fertility and migration, in both areas today's more industrial counties tended in 1900 to have larger percentages of their male population in the productive age group 18–44 ( $.457^*$ ), although, in the Tennessee Valley at least, this tendency had become much weaker since 1860 ( $.626^{**}$ ).<sup>18</sup>

We may therefore conclude that, in the absence of substantial industrial development during 1850–1900, the initial advantages today's more industrial counties enjoyed in terms of superior "original" natural resources and differential windfall gains derived from profitable new cash crops did not suffice to prevent a significant reduction in, or even the elimination of, their initially favorable differentials in per worker farm capital and income. Instead, the indications are that during 1850–1900 there were in both areas long-run equilibrating factor ad-

justments (particularly in labor) of sufficient magnitude to reduce very substantially any intercounty income differences which, had they persisted, might still have been plausibly attributed to differences in "original" resource endowments alone.<sup>19</sup> It is against such a background that the subsequent *divergence* in per worker farm capital and income within each area must be viewed.

#### IV. DIFFERENTIAL INDUSTRIALIZATION AND DIVERGENT AGRICULTURAL INCOMES, 1900–1950

I have already noted that after 1900 both the Tennessee Valley and Piedmont study areas experienced a substantial,

<sup>18</sup> In my analysis of the pre-1900 economic development of the Tennessee Valley, however, I did note certain cultural differences between today's more and less industrial counties which may have persisted throughout the century 1850–1950 and which may have had their roots at least partly in differences in "original" resources. Thus today's more industrial counties had initially enjoyed during 1850–70 an ephemeral industrial development based on local mineral resources that played out, but perhaps not without leaving favorable local attitudes and entrepreneurial talents that encouraged and supported new industries as new opportunities arose after 1900. Similarly, even in their less prosperous days around 1900, today's more advanced counties had lower fertility rates and higher educational investments per school child ( $.564^*$  in 1850,  $.244$  in 1880, but  $.590^{**}$  in 1900), which probably reflected somewhat superior total community resources and, probably more important, a continued belief in the importance of maintaining the "quality" of their population. As a consequence, they probably laid the groundwork for their subsequent industrial rebirth, which was based largely on outside rather than local capital and managerial resources ("Foundations of Economic Development . . .," *op. cit.*, pp. 413–15, 410–11).

It should also be noted that, in the Tennessee Valley area, the windfall gains which wheat brought to today's more advanced counties during 1850–70 were ephemeral: interregional competition eventually took its toll. In the Piedmont area, on the other hand, the initial windfall gains which the introduction of cotton brought to today's more advanced counties were gradually spread (as happened during 1920–40 with burley tobacco in the Tennessee Valley) to even the less advanced counties (*ibid.*, pp. 293–94, 415, and Tang, *op. cit.*, pp. 35–36).

<sup>19</sup> Nicholls, "Some Foundations of Economic Development . . .," *op. cit.*, pp. 415, 405–6, and Tang, *op. cit.*, pp. 88, 171.

but highly uneven, degree of industrial-urban development and that during the same period intercounty differentials in per worker farm capital and income sharply increased. As Table 1 shows, while today's more industrial counties did not in 1900 have significantly higher farm income per worker (.185) than their less industrial neighboring counties, by 1940 they clearly did so (.678\*\*), and in 1954 this relationship stood at its highest level (.744\*\*) in history. For either study area, the agriculture of today's more advanced counties—after starting on a nearly equal footing in farm income per worker in 1900—had forged increasingly far ahead of the agriculture of those neighboring counties which had not enjoyed comparable industrial-urban development during the subsequent fifty-year period.<sup>20</sup>

Can we attribute these persistent and increasing intercounty differentials in per worker farm income after 1900 to differences in county rates of industrial-urban development? Tang and I have both concluded that we can. Despite the fact that the period 1900–1950 saw a quickening of factor mobility, such resource adjustments no longer appear to have been as effective in dampening existing agricultural income differentials between counties as they had been in the pre-industrial period. Apparently, the dynamic nature of industrial-urban development had a disequilibrating effect on farm incomes per worker which far more than offset the equilibrating effect of continued factor transfers during the more recent period of industrialization. Presumably, the reason was that local industrial-urban development transmits its effects on local agricultural productivity and incomes through its impact on local factor and product markets, which function more efficiently the greater the level of

nearby industrial-urban development. Let us examine this basic hypothesis further, first with reference to the capital market and then with reference to the labor market.

*The impact of industrialization on the local capital market.*—One would expect that those counties which have since 1900 enjoyed differentially greater industrial-urban development would, because of an initial influx of outside capital (most new industry was so financed) and subsequent growth of local employment and income, have shown increasingly greater bank deposits per capita as compared with those of the less developed counties. In both the Tennessee Valley and the Piedmont, today's more advanced counties had as early as 1900–1905 significantly larger per capita (primary) bank deposits. For the Tennessee Valley, this relationship was highly significant throughout 1900–1954 but slightly weakened during the half-century (.809\*\* in 1900, .714\*\* in 1950) as the number and dispersal of banks in the study area grew (Table 2). However, as in the Piedmont area, the absolute differences in per capita bank deposits (in constant prices) increased enormously as between more and less industrial counties of the Tennessee Valley (for example, in 1926 prices, the interquartile range of per capita bank deposits for the twenty Tennessee Valley counties increased from \$15 to \$135 during 1900–1950). As a consequence, the increasing advantage of the more rapidly industrializing counties in terms of volume of bank deposits was presumably reflected, through the greater availability of agricultural credit, in a higher rate of capital formation in the nearby agriculture.<sup>21</sup>

<sup>21</sup> *Ibid.*, pp. 113–17. Tang went on to demonstrate (pp. 118–22 and 124–25) that, per farm worker, both short- and long-term bank loans to agriculture were significantly greater in the more industrial counties than in the less industrial counties of the Piedmont.

<sup>20</sup> Tang, *op. cit.*, pp. 66–72.

Of course, insofar as today's more industrial counties had higher rates of capital formation in their agricultural sector, these higher rates might be in part attributable to inherently superior land resources. For example, the Tennessee Valley's more developed counties already had in 1900 significantly higher per acre farm-land values (.627\*\*). Furthermore, throughout 1900-1954, they maintained a strong relative advantage over the less developed counties in per acre yields of wheat, although the relative advantage in corn and oat yields which they enjoyed during most of 1900-1940 had been

eliminated or substantially weakened by 1954 (Table 2). These latter data suggest that, unlike in the Piedmont area,<sup>22</sup> those counties enjoying higher rates of industrial-urban development after 1900 at best maintained, rather than strengthened, their favorable position in terms of crop yields. On the other hand, the Tennessee Valley's more industrial counties did strengthen their already strong tendency to have higher per acre farm-land

<sup>22</sup> *Ibid.*, pp. 142-44. In the Piedmont, Tang found that during 1900-1940 the industrial counties gained sharply on their less industrial neighbors in per acre yields of cotton, wheat, corn, and oats.

TABLE 2

COEFFICIENTS OF RANK CORRELATION BETWEEN PER CAPITA NON-AGRICULTURAL PAYROLLS, 1954, AND SELECTED INDEXES RELATED TO LOCAL CAPITAL MARKET, UPPER EAST TENNESSEE VALLEY STUDY AREA, SELECTED YEARS, 1900-1954\*

| Item   | 1900   | 1910   | 1920   | 1930   | 1940    | 1950                | 1954    |
|--|--------|--------|--------|--------|---------|---------------------|---------|
| Per capita bank deposits.....  | .809** | .805** | .826** | .779** | .746**  | .729**              | .714**  |
| Per acre value of all farmland.....                                    | .627** | .561** | .722** | .805** | .773**  | .773**              | .792**  |
| Per cent of all farmland improved....                                  | .233   | .255   | .272   | .242   | .549**  | .516*               | N.a.    |
| Per acre yield:  |        |        |        |        |         |                     |         |
| Corn.....  | -.071  | .492*  | .395*  | .441*  | .426*   | .226                | .063    |
| Wheat.....   | .701** | .486*  | .593** | .577** | .763**  | .529*               | .787**  |
| Oats.....  | .542** | .550** | .460*  | .456*  | .543**  | .503*               | .320    |
| Tobacco.....   | -.030  | .161   | .256   | -.308  | .247    | -.051               | -.236   |
| Per cent of gross cash farm sales:                                     |        |        |        |        |         |                     |         |
| Tobacco.....   | N.a.   | N.a.   | N.a.   | -.072  | -.418*  | -.603**             | -.632** |
| Livestock.....   | N.a.   | N.a.   | N.a.   | -.004  | .312    | .580**              | .621**  |
| Per farm acreage of all farmland.....                                  | .282   | .089   | -.161  | -.233  | -.242   | -.234               | -.198   |
| Per farm total value of capital.....                                   | .553** | .441*  | .504*  | .633** | .448*   | .559** <sup>b</sup> | .651**  |
| Per farm man-years of labor input....                                  | .582** | .501*  | .414*  | -.158  | -.609** | -.316 <sup>b</sup>  | -.386*  |
| Index of total inputs per farm.....                                    | .673** | .546** | .541** | .359   | -.028   | .198 <sup>b</sup>   | .191    |
| Per farm gross farm income.....  | .283   | .463*  | .388*  | .418*  | .298    | .382 <sup>b</sup>   | .304    |
| Per farm net farm income.....  | N.a.   | N.a.   | N.a.   | N.a.   | .147    | .182 <sup>b</sup>   | .039    |
| Per worker total value of farm capital                                 | .344   | .387*  | .484*  | .663** | .808**  | .798** <sup>b</sup> | .841**  |
| Instructional expenditures per school child.....                       | .590** | .559** | .719** | .719** | .740**  | .699**              | .714**  |
| Per cent of children 16-17 years old attending school.....             | N.a.   | -.204  | -.406* | -.134  | .585**  | .689**              | N.a.    |
| Median years of school completed by persons 25 years old and over..... | N.a.   | N.a.   | N.a.   | N.a.   | .790**  | .820**              | N.a.    |

\* Coefficients of correlation between county ranks in the particular index in each selected year and county ranks in 1954 per capita non-agricultural payroll.

<sup>b</sup> The coefficients for commercial farms and part-time farms separately were:

|                                     | Commercial | Part-Time |
|-------------------------------------|------------|-----------|
| Per farm capital.....               | .627**     | .812**    |
| Per farm labor.....                 | .281       | -.674**   |
| Index of total inputs per farm..... | .540**     | .100      |
| Per farm gross income.....          | .479*      | .206      |
| Per farm net income.....            | .392*      | -.064     |
| Per worker farm capital.....        | .671**     | .800**    |

\* Significant to 5 per cent level.

\*\* Significant to 1 per cent level.

values between 1900 (.627\*\*) and 1954 (.792\*\*); and their initially weak tendency (.233 in 1900) to have larger percentages of "improved" farm land had become significant by 1940-50 (.549\*\* and .516\*). Therefore, to a limited extent at least, the more industrial counties were apparently able to increase their rate of investment in their farm land more than the less industrial counties found possible. While this fact was probably in part due to the greater ability of their farmers to finance their own improvements because of a superior resource base, we have already seen that such superiority was not sufficient in 1900 to give them a significant advantage in per worker farm income—the crucial variable so far as internal savings and capital formation are concerned. Hence, as in the Piedmont, their increasing rate of investment in their farm land was primarily attributable to the greater availability of farm credit which accompanied nearby industrial-urban development and to the consequent more perfect functioning of local capital markets.

It is also worth noting that industrial-urban development was in both areas associated with new enterprise combinations in agriculture in which livestock products played an increasingly important part. In the Tennessee Valley, today's more advanced counties did not derive larger proportions of their gross cash farm income from livestock in 1930 (— .004) but showed a very strong tendency to do so by 1954 (.621\*\*). During the same period, the more advanced counties—which in 1930 had not differed significantly in terms of the proportion of their gross cash farm income derived from the major staple cash crop, tobacco (— .072)—had by 1954 become relatively far less dependent upon tobacco than the less industrial counties (— .632\*\*). These

sharp shifts in the relative importance of different farm products in the more and less industrial counties undoubtedly reflect not only (a) an increasing relative farm-labor scarcity (via improved labor markets) unfavorable to a labor-intensive crop and (b) the improved local markets for livestock products created by nearby industrial-urban development, but also (c) concomitant improvement in the functioning of local capital markets by which expansion of livestock enterprises might be financed.<sup>23</sup>

What was the influence of industrial-urban development on the resources (inputs) and income (output) of the average farm (Table 2)? In the Tennessee Valley, at no time during 1900-1954 did the average farms of the more developed counties have significantly different acreages of all farm land than those of the less developed counties. Indeed, over the fifty-year period, there was some tendency (primarily due to their more rapid increase in part-time farms) for the more developed counties to have average farms with *less* acreage than the average farms of the less developed counties (the coefficient was .282 in 1900 and — .234 in 1950). Furthermore, largely because of the persistently higher land values in the more advanced counties, their average farms had throughout 1900-1954 significantly greater total farm capital than those of the less advanced counties. While this tendency weakened somewhat between 1900 (.553\*\*) and 1940 (.448\*), by 1954 it had reached its highest level (.651\*\*). On the other hand, throughout 1900-1954, the average farms of the more developed counties had significantly different man-years of labor inputs from

<sup>23</sup> *Ibid.*, pp. 104-9. Unlike in the Tennessee Valley, however, in the Piedmont the major cash crop (cotton rather than tobacco) as late as 1940 was not of significantly less relative importance in the more than in the less industrial counties.

those of the less developed counties. In the earlier years they showed a strong tendency to have *more* labor inputs (.582\*\* in 1900), but from 1940 (— .609\*\*) on, the tendency was reversed due to the increasing relative importance of part-time farms in the averages. (For commercial farms only, the relationship in 1950 was still positive at .281.)

To what extent are these contrary trends indicative of mere factor substitution without scale adjustments? In other words, did the average farms of the more industrial counties simply use greater capital inputs to offset fewer labor inputs, without increasing their total inputs or value of output? To answer this question on the input side, I devised an index of total inputs (which may be viewed as a crude index of scale), based on the ranking of the sum of ranks of each county's per farm capital and labor inputs. According to this index, the average farms of today's more industrial counties enjoyed a strong relative scale advantage in 1900 (.673\*\*) which, as industrialization proceeded with a concomitant increase in numbers of part-time farms, had disappeared by 1940 (— .028). Thereafter, there was a tendency for the relationship to be restored, with the average *commercial* farms of the more industrial counties in 1950 employing significantly larger total inputs than their counterparts in the less industrial counties (.540\*\*).<sup>24</sup>

Alternatively, we may measure farm size by total value of output instead of total inputs. If we do, we find that, relative to the average farms of today's less industrial counties, the average farms of

the more industrial counties showed a moderate tendency during 1900–1954 to have larger gross farm incomes. However, this tendency did not become markedly stronger as industrialization proceeded. Again, however, the reason would appear to be primarily the result of the depressing effects of the growing number of part-time farms in the industrial counties upon per farm averages. Thus, in 1950 (the only year in which commercial and part-time farms can be separately treated), the average *commercial* farms of the more industrial counties had significantly larger gross farm incomes than those of the less industrial counties (.479\*), whereas the average part-time farms did not (.206). These coefficients may be compared with .283 for *all* farms in 1900.

Nonetheless, it is evident that, whether measured in terms of total inputs or total output, our indexes of average farm size (scale) are rather imperfect and are relatively insensitive to the effects of differential industrial development. The value of farm capital per farm worker, on the other hand, reflects very fully these effects. Thus, as we saw in Table 1, in 1900 today's more industrial counties did not have a significant larger amount of farm capital per farm worker than did today's less industrial counties (.344). By 1940, however, they showed a very strong tendency to do so (.808\*\*) and maintained their advantage through 1954 (.841\*\*). These indexes strongly support the view that industrial-urban development was highly associated with relatively high capital-labor ratios in the agriculture of the Tennessee Valley as in Piedmont agriculture.<sup>25</sup> However, for both areas the weight of the evidence suggests that the strengthening of this association was much more the result of

<sup>24</sup> For a more detailed treatment of this matter of total inputs, see my most recent articles, "Industrial-Urban Development and Agricultural Adjustments . . .," *op. cit.*, pp. 139 and 141–42, and "Factors Affecting Gross Farm Income . . .," *op. cit.*, pp. 358–62.

<sup>25</sup> Tang, *op. cit.*, pp. 126–28.

differentially greater reductions in the more industrial counties' farm-labor inputs (the denominator of the capital-labor ratio) than to differentially greater increases in their farm capital (the numerator).<sup>26</sup> Hence, the relative impact of industrial-urban development on the efficiency of local factor markets appears to have been greater in the labor market than in the capital market.

Under such circumstances, a comparison of the responses of commercial and part-time agriculture (possible only in 1950) is instructive. In his Piedmont study, Tang concluded that the part-time sector was far more responsive to the positive income effect of local industrial development. This difference he attributed (a) to the greater ease in the more industrial counties of transferring underemployed farm labor to other employment without appreciably changing total capital or output, part-time farms by definition being the principal beneficiaries; and (b) to the slowness and difficulty with which the commercial farms, even in the more developed counties, were able to acquire additional land and other capital that might have justified the retention in agriculture of larger labor forces per farm. In the Tennessee Valley area, however, I found that the commercial farms responded favorably to local industrial development at least as well as neighboring part-time farms in terms of farm capital per worker, and considerably better in terms of total inputs and income and per worker net farm income. The stronger response of commercial farms in the Tennessee Valley, as compared with commercial farms in the Piedmont, was, I concluded, due to a larger proportion of its commercial-

farm capital being agriculturally productive, to its less imperfect capital markets, and to its greater availability of superior managerial talent.<sup>27</sup>

Finally, on the capital side, today's more industrial counties in both areas have shown higher rates of investment in human beings than have their less industrial neighbors. In the Tennessee Valley area, educational outlays per school child were relatively greater in today's more industrial counties throughout the century 1850-1950. However, between 1900 (.590\*\*) and 1920 (.719\*\*), this tendency was strengthened somewhat and has been maintained to date. School attendance of children aged 16-17 in the more and less advanced counties was not significantly different in 1910 (-.204) but by 1920 was significantly less (-.406\*) in today's more advanced counties. However, these initially adverse effects of industrial development had been fully reversed by 1950, when the more industrial counties had a very strong relative advantage over their less industrial neighbors in both high-school attendance (.689\*\*) and median years of school completed by persons of 25 years of age or older (.820\*\*). Thus, as in the Piedmont<sup>28</sup> we find that industrial-urban development has been associated with relatively greater capital investment in the human agent—a factor clearly favorable to further economic development through its positive influence on both labor productivity and awareness of economic opportunities.

Our findings on the capital market may be summarized as follows. Local industrial-urban development, typically made possible by an influx of outside

<sup>26</sup> *Ibid.*, pp. 197-98; Nicholls, "Effects of Industrial Development . . .," *op. cit.*, p. 1648. See below, p. 334.

<sup>27</sup> Nicholls, "Industrial-Urban Development and Agricultural Adjustments . . .," *op. cit.*, pp. 144-49. See below, pp. 334-36.

<sup>28</sup> Tang, *op. cit.*, pp. 138-41.

non-agricultural capital, increases local personal incomes and savings, and by increasing the total resources of local banking and credit institutions redounds to the benefit of the nearby agriculture. Agriculture benefits directly from the consequent greater availability of capital which facilitates investments in land improvement, the development of more profitable capital-intensive enterprises (particularly livestock), the raising of capital-labor ratios, and often an increased scale of farming operations as well—all of which increase farm incomes per worker relative to those in the less industrial-urban neighboring counties. Agriculture also benefits indirectly from the effects of more adequate credit on the development of local firms serving agriculture as suppliers of farm-production goods and as purchasers of farm products; and from the effects of an expanding tax base on the supply and quality of public services (education, roads, health, etc.) available to nearby rural people.

*The impact of industrialization on the local labor market.*—In 1900, today's more developed counties in the Tennessee Valley did not have a significantly smaller percentage of rural population than today's less developed counties ( $-.162$ ), but from 1920 on they showed a very strong tendency ( $-.884^{**}$  in 1950) to do so (Table 3). Even in 1900, today's more developed counties strongly tended to have larger percentages of their total work force engaged in manufacturing ( $.635^{**}$ ) and smaller percentages in agriculture ( $-.808^{**}$ ). Thus these particular indexes in 1900 had excellent predictive value with reference to the relative economic development of the several counties in the study area during the next half-century. Nonetheless, inter-county differences in these indexes were too small in 1900 to have much differen-

tial effect on the nearby agriculture. During 1900–1950, however, the interquartile range in percentage of manufacturing employment increased from 3.7 to 13.4 and the corresponding range in percentage of agricultural employment increased from 10.6 to 29.1. With these rapidly growing intercounty differences in relative non-agricultural employment, the counties enjoying the most industrial-urban development have also in recent decades greatly increased their per worker gross farm income relative to the counties which lagged behind. For the latter index, the interquartile range (in constant 1926 dollars) increased from \$54 to \$259 during 1900–1950 and, by 1954, the tendency for today's more industrial counties to have higher per worker gross farm incomes was very strong ( $.744^{**}$ ). The same was true for rural levels of living ( $.780^{**}$ ).

By 1940, annual average earnings per worker in manufacturing and retail trade in the Tennessee Valley were \$899 and \$759, respectively, as compared with \$440 gross farm income per worker in agriculture. Since the latter figure includes net returns to all factors of production in agriculture rather than labor returns only, an attempt was also made to estimate for 1940 the *residual* returns to the labor factor in agriculture.<sup>29</sup> For the twenty Tennessee Valley counties, the average net labor return per worker in agriculture was \$270—less than one-third of the annual earnings per non-agricultural worker. Within the Tennessee Valley (as in the Piedmont), today's more advanced counties in 1940 enjoyed relatively favorable labor returns in the non-agricultural industries and agriculture alike. For per worker annual earnings in manufacturing and retail trade,

<sup>29</sup> The method used was the same as Tang's (*ibid.*, pp. 129–36).



the coefficients in 1940 were .629\*\* (as compared with only -.056 in 1900) and .652\*\*, respectively. For per worker gross farm income, the relationship was even stronger (.678\*\*) but for per worker residual labor returns in agriculture the coefficient was considerably smaller although still significant (.414\*).

It is worth noting that this relationship between net labor returns per farm worker and industrial-urban develop-

ment was substantially weaker in the Tennessee Valley than in the Piedmont. Furthermore, for 1940 my estimates of net returns on all capital (land, service buildings, and non-real estate inventories) per \$100 of capital indicated that in the Tennessee Valley the more industrial counties did not earn significantly different returns on their productive farm capital (.069) than did the less industrial counties. In the Piedmont, on

TABLE 3  
COEFFICIENTS OF RANK CORRELATION BETWEEN PER CAPITA NON-AGRICULTURAL PAYROLLS, 1954,  
AND SELECTED INDEXES RELATED TO LOCAL LABOR MARKET, UPPER EAST TENNESSEE  
VALLEY STUDY AREA, SELECTED YEARS, 1900-1954<sup>a</sup>

| Item  | 1900                 | 1910    | 1920    | 1930    | 1940                | 1950                | 1954    |
|---|----------------------|---------|---------|---------|---------------------|---------------------|---------|
| Per cent of population:   |                      |         |         |         |                     |                     |         |
| Rural.....  | -.162                | -.352   | -.627** | -.678** | -.553**             | -.884**             | N.a.    |
| Rural-farm.....   | N.a.                 | N.a.    | N.a.    | -.773** | -.816**             | -.808**             | -.828** |
| Per cent of work force employed:  |                      |         |         |         |                     |                     |         |
| Agriculture.....  | -.808**              | -.816** | -.792** | -.785** | -.806**             | -.816**             | N.a.    |
| Manufacturing.....  | .635**               | N.a.    | .852**  | .869**  | .836**              | .878**              | N.a.    |
| Per worker average earnings:  |                      |         |         |         |                     |                     |         |
| Manufacturing.....  | -.056                | N.a.    | .344    | .423*   | .629**              | .802**              | .802**  |
| Retail trade.....   | N.a.                 | N.a.    | N.a.    | .356    | .602**              | .622**              | .531*   |
| Per worker gross farm income.....   | .185                 | .313    | .186    | .438*   | .678**              | .624** <sup>b</sup> | .744**  |
| Per worker net farm income.....   | N.a.                 | N.a.    | N.a.    | N.a.    | .586** <sup>b</sup> | .633** <sup>b</sup> | .375    |
| Rural level-of-living index.....  | N.a.                 | N.a.    | N.a.    | .564**  | .705**              | .729**              | .780**  |
| Crude rate of natural increase.....   | -.456** <sup>c</sup> | N.a.    | -.267   | -.669** | -.578**             | .007                | N.a.    |
| Fertility rate.....   | -.426** <sup>c</sup> | N.a.    | N.a.    | -.663** | -.761**             | -.606**             | N.a.    |
| Average size of family.....   | -.484*               | -.292   | -.382*  | -.562** | -.468*              | -.744**             | N.a.    |
| Net out migration rate, previous decade.....  | -.453*               | -.511*  | -.608   | -.783** | -.528*              | -.902**             | N.a.    |
| Per cent of male population aged 18-44.....   | .457*                | N.a.    | .418*   | .879**  | .828**              | .851**              | N.a.    |
| No. males per 100 females, aged 18-44.....  | N.a.                 | N.a.    | -.120   | .054    | -.575**             | -.708**             | N.a.    |
| Per cent of total farm-operator man-years worked off farm.....                                    | N.a.                 | N.a.    | N.a.    | N.a.    | .402*               | .599**              | .662**  |
| Per cent of farms on all-weather roads.....   | N.a.                 | N.a.    | N.a.    | .784**  | .325                | .391*               | N.a.    |
| Per cent of gross farm income from dairy, poultry, vegetables, and horticultural specialties..... | N.a.                 | N.a.    | N.a.    | N.a.    | .390*               | .481*               | .558**  |

<sup>a</sup> Coefficients of correlation between county ranks in the particular index in each selected year and county ranks in 1954 per capita non-agricultural payroll.

<sup>b</sup> The coefficients for commercial farms and part-time farms separately were:

|                                   | Commercial | Part-Time |
|-----------------------------------|------------|-----------|
| Per worker gross farm income..... | .525*      | .738**    |
| Per worker net farm income.....   | .444*      | .531*     |

Other important indexes produced the following coefficients for all farms: residual (net) labor returns per worker in agriculture 1940, .414\*; and median net cash incomes (all sources), rural-farm families 1950, .847\*\*.

<sup>c</sup> Data for 1880 rather than 1900.

\* Significant to 5 per cent level.

\*\* Significant to 1 per cent level.

the other hand, the more industrial counties strongly tended to have *lower* returns on their farm capital than did their capital-poor, less industrial neighbors. While our estimates of net factor returns are very crude, these inter-area differences support the view that the Tennessee Valley had much less imperfect capital markets and somewhat less imperfect labor markets than did the Piedmont.<sup>30</sup>

In the previous section, we nonetheless saw that the agriculture of the more industrial counties in both areas enjoyed a higher rate of capital formation and an increasingly favorable ratio of farm capital to farm labor as compared with the agriculture of the less industrial counties. In the present section, we have just seen that the agriculture of the more industrial counties also enjoyed higher net returns to labor. If local labor markets had functioned perfectly, however, the direction and magnitude of movements of agricultural labor would have been such as to offset these differential rates of capital formation while narrowing, if not eliminating, the gap in net labor returns in agriculture. To what extent, and why, did local labor markets in the two study areas function less than perfectly?

In any given county, the farm-labor force is a joint function of demographic characteristics (particularly fertility rates) and changes of residence and/or occupation. Let us look at the matter of demographic characteristics first (Table 3). In recent decades, as one would expect, the Tennessee Valley's more industrial-urban counties had significantly lower crude rates of natural increase, lower fertility rates ( $-.606^{**}$  in 1950), and smaller average family sizes ( $-.744^{**}$ ) than did its less developed counties. Such relationships might be expected to be the consequence of indus-

trial-urban development. It is important to emphasize, however, that these inter-county differences clearly antedated the area's modern period of industrial-urban development, hence may have been in considerable part cause rather than effect of later differences in rates of economic progress.<sup>31</sup>

Because through higher fertility rates they historically kept the inflow of new persons into their population reservoirs at higher levels, today's less industrial counties would have had to have higher outmigration rates than today's more advanced counties, even if the latter had not experienced any significant industrial-urban growth. Indeed, this appears to have happened during 1850-1900, with today's more industrial counties having had sufficiently smaller outmigration rates (the coefficient for 1890-1900 was  $-.453^{*}$ ) so that in 1900 they did not have significantly greater per worker farm capital (.344) or per worker farm income (.185). Thus, in the pre-industrial period, local labor markets, though imperfect, apparently sufficed to eliminate substantial intercounty differences in agricultural productivity and income. As today's more advanced counties pulled ahead industrially during 1900-1950, however, increasingly great inter-county differences in outmigration rates were not enough to prevent their agriculture from also pulling ahead in per-worker capital and income. For example, throughout 1900-1950, today's more industrial counties had significantly lower outmigration rates per decade—a relationship that never fell below  $-.51^{*}$  (1900-1910) and reached  $-.902^{**}$  during 1940-50. Since it was not unusual for the least industrial third of the twenty

<sup>31</sup> See n. 18. For a more detailed analysis of population characteristics, see my article, "Human Resources and Industrial Development . . .," *op. cit.*, pp. 300 ff.

<sup>30</sup> *Ibid.*, p. 133, and above, p. 331.

counties to lose over 20 per cent of their total populations per decade by net outmigration, it should be clear that outmigration alone can hardly be expected to bring about a general equilibrium in the area's agricultural productivity and income.

Since we have here been dealing with *county* rates of outmigration, however, the question still remains as to how much of the migration was internal to the study area and how much was from the area as a whole to other regions. That much was internal is shown by the fact that today's more industrial counties showed an increasingly strong tendency to have larger proportions of their male populations in the most productive age group 18 through 44. This tendency, unlike in the Piedmont,<sup>32</sup> was already significant in 1900 (.457\*). By 1940, however, it was highly significant in both study areas (.828\*\* in the Tennessee Valley). That the more industrial counties have had larger proportions of their male populations in the more productive years is, of course, a very favorable situation, reflecting not only their lower fertility rates but the age-selectivity of their inmigration.

However, many outmigrants from individual counties clearly moved beyond the study area to other regions in which economic opportunities were better. For example, in every decade during 1900-1950, the Tennessee Valley study area as a whole lost from 5 to 14 per cent of its population by migration to other areas. That increasingly these outmigrants from the study area were disproportionately male is shown by the fact that since 1940 today's more industrial coun-

ties in both the Tennessee Valley and the Piedmont<sup>33</sup> had *fewer* males per hundred females in the age group 18-44 (-.708\*\* in 1950)—a difference which can be explained only by sex selection in migration patterns, with females more frequently migrating shorter distances from less to more industrial counties within the study area and males more frequently migrating out of the area as a whole.

In this connection, it is interesting to note that during 1940-50 both the Tennessee Valley and the Piedmont had increases in total real farm capital (in constant 1939 dollars) of 10 per cent but that the decline in farm-labor input was substantially less in the Tennessee Valley (16 per cent) than in the Piedmont (35 per cent). Since during 1940-50 the Tennessee Valley area had a net outmigration rate (12 per cent) almost twice that of the Piedmont (7 per cent), it would appear that the reduction in the farm-labor force in the Tennessee Valley was to a much less extent the result of shifts to local non-farm employment within the area than of outmigration from the study area. While in both areas the counties less developed *at the beginning of the decade 1940-50* subsequently had significantly higher net outmigration rates (-.813\*\* and -.641\*\*, respectively), in the Tennessee Valley they also made relatively larger downward adjustments in their farm-labor force than did the more developed counties (-.483\*), whereas in the Piedmont the initially less developed counties at best matched those of the more developed counties (.034) in reducing the farm-labor force.<sup>34</sup>

Despite such differences between the two study areas, it is clear that in both

<sup>32</sup> Tang, *op. cit.*, p. 171. Unlike in the Tennessee Valley, net outmigration rates in the Piedmont were not significantly different in today's more and less industrial counties during 1900-1920 although they were thereafter (*ibid.*, p. 166).

<sup>33</sup> *Ibid.*, p. 173.

<sup>34</sup> See my article, "Industrial-Urban Development and Agricultural Adjustments . . ." *op. cit.*, pp. 144, 141, 138.

areas the magnitudes of migration rates and reductions in the farm-labor force were very large. Nonetheless, in neither area was the outmigration of farm people (and the influx of farm capital) sufficient to raise returns to the human agent in the agriculture of the less industrial counties to a level comparable with that earned in the agriculture of the more industrial counties. Since both areas continued to be deficit areas in terms of non-agricultural employment, nearby job opportunities have been "rationed," in the sense that more farm people would shift to local non-farm jobs at prevailing wages if these jobs were available. In this rationing process, industrial employers can find numerous rational reasons (apart from emotional ties) for giving local residents some preference over long-distance commuters. Furthermore, the time and cost of commuting is in part a function of physical distance, a factor which is compounded by the typically poorer rural roads in the less developed counties. As late as 1950, today's more industrial counties in the Tennessee Valley had significantly larger percentages of all farms on all-weather roads (.391\*), although their relative advantage had sharply declined since 1930 (.784\*\*).<sup>35</sup>

Of course, as an alternative to commuting long distances, farmers in the less developed counties can move to the more developed counties. But changes of occupation are much more difficult to effect if they also require changes of residence. Farm operators of the more industrial counties have found it possible to work relatively more of their time in off-farm work than those in the less developed counties (.402\* in 1940, .662\*\* in 1954). For such nearby farm operators, not only are the opportunities for off-farm work greater but the choice of off-farm em-

ployment (since it involves no change of residence) is a much easier one to make, particularly in an area (such as the Tennessee Valley) where most farms are owner-operated. Despite his stronger attachment to the land, the owner-operator of a small farm in a more industrial county can rather readily solve his problem of underemployment (as can the members of his family) by shifting part of his labor to non-farm work—which may also supply an excellent source for self-financing of farm capital improvements—thereby raising his labor productivity and income. Thus, unlike his counterpart in a less developed county, he can keep his small acreage and raise his combined income too. On the other hand, the farm tenant is not always as mobile as his more tenuous attachment to the land might suggest. This is especially true where (as in the Piedmont) not only are tenancy rates high but most tenants are Negro sharecroppers, because landlords commonly strongly discourage sharecropper families from engaging in off-farm work. In addition, if they are Negroes, they may have to resort to long-distance migration beyond the area if they are to overcome their agricultural underemployment.<sup>36</sup>

In spite of such impediments to farm-labor mobility as high non-economic valuations placed on land ownership and the rigidities imposed by landlord-tenant relationships and race, it is clear that superior non-farm employment opportunities in the more developed counties make for significantly more efficient labor mar-

<sup>35</sup> *Ibid.*, pp. 162, 158, 136-38, 177-90. Because of the much greater importance of tenancy and Negroes in his study area, Tang has given much more attention to the effects of both than I needed to. See also his recent article, "Economic Development and Changing Consequences of Race Discrimination in Southern Agriculture," *Journal of Farm Economics*, LI (December, 1959), 1113-26.

<sup>36</sup> Tang, *op. cit.*, pp. 100, 156-58.

kets. As a consequence, local industrialization increases the actual or imputed cost of the human agent in agriculture toward (or, in the absence of job rationing, to) the prevailing level of non-farm wages, in turn forcing those who remain in agriculture to reorganize their farms to raise labor productivity enough to cover the higher labor cost. Such a reorganization may take one of two forms.

First, as we have just seen, nearby uneconomically small, full-time farms can become part-time farms without any increase in the amount of farm land or other capital (although self-financing of farm capital is also facilitated) if one or more members of the farm family takes off-farm employment. This adjustment, which is obviously easiest to effect, will tend to raise the productivity of the residual farm labor to the level in alternative non-farm employment. Small full-time farmers in the less industrial counties, on the other hand, lack both equal opportunities and comparable economic pressures to raise their productivity and incomes by part-time farming. Thus in both the Tennessee Valley and the Piedmont the part-time farms of the more industrial counties had in 1950 larger per worker farm capital (.800\*\*), per worker gross farm income (.738\*\*), and per worker net farm income (.531\*) than did the part-time farms of the less developed counties (Table 2, nn. *a* and *b*; Table 3, nn. *a*, *b*, and *c*). In the Tennessee Valley the part-time farms of the more industrial counties strongly tended to have relatively larger capital inputs (.800\*\*) per farm and, in both areas, relatively smaller labor inputs (−.674\*\*) per farm than did those of the less developed counties, but these differences were almost completely offsetting as far as average farm scale was concerned (Table 2, nn. *a*

and *b*). For example, the part-time farms of the more and less industrial counties did not have significantly different *total* inputs (.100), gross farm income (.206), or net farm income (−.064) per farm. These findings indicate that, as far as the part-time sector was concerned, the more industrial counties of each study area enjoyed a strong advantage not only in farm-labor productivity but in the ratio of farm capital to farm labor—with the Tennessee Valley experiencing larger capital adjustments and smaller labor adjustments than the Piedmont—but neither area made significant scale adjustments in its part-time farms.<sup>37</sup>

The second type of economic reorganization of farm firms resulting from local industrialization and its effects on farm-labor costs (and the availability of farm capital) is the enlargement of full-time farms, by which farm-labor productivity might again be raised enough to cover the increased opportunity cost of labor. In many instances, it may be possible to reorganize such full-time farms without enlargement of acreage or reduction of the farm-labor force. Since farm-land values may tend to be relatively high (although for quite different reasons) in both the more and less developed counties, a solution involving more intensive land use will presumably be favored insofar as feasible in either setting.

However, in the less developed counties, the possibilities of intensive land use in any case will probably have been exhausted. Lacking both alternative local non-farm employment opportunities and adequate rates of intercounty labor and capital mobility, these less developed counties have large numbers of farm workers the alternative use value of

<sup>37</sup> Tang, *op. cit.*, pp. 160–61, 204–7; and my article "Industrial-Urban Development and Agricultural Adjustments . . .," *op. cit.*, pp. 146–47.

whose labor locally approaches zero. As a consequence, they have probably bid up the local price of farm land to levels which produce no more than a very low residual contribution of labor to agricultural output. At the same time, they have probably concentrated on labor-intensive products (for example, tobacco or cotton) which, while using land intensively, require little non-real estate capital, hence are associated with low farm-labor productivity. As we have already observed (Table 2), the less developed counties have shown an increasingly strong tendency to derive relatively larger shares of their gross cash farm income from tobacco ( $-.632^{**}$  in 1950).

In the more developed counties, farmland values may also be high because of the greater competition for land for non-agricultural uses, again favoring farm enterprises which demand intensive land use. However, full-time farms in such counties will probably have access to more non-real estate capital and more favorable local product markets, permitting concentration on other intensive farm products which will enable given farm labor to be more fully utilized while raising the productivity of labor to cover its locally higher opportunity cost. For example, it is significant in this connection (Table 3) that in both study areas the more industrial counties derived significantly larger percentages of their gross farm incomes (cash and kind) from dairy and poultry products, vegetables, and horticultural specialties between 1940 (.390\*) and 1954 (.558\*\*).<sup>18</sup> Being perishable, these products are probably more sensitive to local industrial-urban development than most other products.

Finally, full-time farms may be reorganized by an increasing emphasis on farm enterprises (for example, hay and

beef cattle) which make less intensive use of labor and more extensive use of land. In this case, a reduced farm-labor force and increased acreage (through consolidation of small farms) are required. Such adjustments are very difficult to make in the less developed counties unless (as rarely happens) the rate of out-migration is sufficiently high to reduce substantially the pressure of the population on the local farm-land supply, farmland values falling to levels which make economic the consolidation of the land by remaining farmers. Even then, lacking adequate access to capital, such community-wide economic reorganization of agriculture may not be possible in the less industrial counties. Even in the more industrial counties—despite more favorable capital, labor, and product markets—economic reorganization involving land consolidation faces considerable difficulties in view of increasing non-agricultural site values of farm land and the peculiar institutional barriers to the purchase of contiguous parcels of land. Nonetheless, one would expect a tendency toward this type of reorganization of full-time farms, particularly in the more outlying agricultural areas of the more industrial counties.

In any case, with or without land consolidation, the full-time (commercial) farms of the Tennessee Valley's more industrial counties had in 1950 larger per worker farm capital (.671\*\*), per worker gross farm income (.525\*), and per worker net farm income (.444\*) than did those of its less developed counties (Table 2, nn. *a* and *b*; Table 3, nn. *a*, *b*, and *c*). While the Tennessee Valley's more industrial counties did not have significantly larger labor inputs per commercial farm (.281), they enjoyed substantially larger capital inputs per commercial farm (.627\*\*). As compared with

<sup>18</sup> Tang, *op. cit.*, p. 105.



the commercial farms of the less developed counties, those of the more advanced counties showed larger "scale," whether measured by total inputs (.540\*\*), gross farm income (.479\*), or net farm income (.444\*) per farm. The commercial-farm sector of the Piedmont showed similar but somewhat weaker relationships. Hence, the commercial (full-time) farms of both areas responded significantly to local industrial-urban developments. However, their response was somewhat greater in the Tennessee Valley, not only in terms of somewhat greater "scale" adjustments but through far greater increases in the ratio of capital to labor, with much more substantial gains in farm-labor productivity as a result. While the more industrial counties of each study area enjoyed more efficient capital and labor markets, those of the Tennessee Valley showed the greater advantage, perhaps because they had relatively fewer farm tenants, Negroes, and female manufacturing employees than the Piedmont.<sup>39</sup>

Elsewhere, I have shown that there was a much weaker tendency in the Tennessee Valley (.378\*) than in the Piedmont (.540\*\*) for per worker net farm income on part-time farms to be a larger percentage of per worker income on commercial farms, the more industrialized the county. Despite this evidence of the greater relative inefficiency of part-time farming in the Tennessee Valley, the inclusion of non-farm income presents a much more favorable picture. Unlike in the Piedmont, per capita non-agricultural payrolls (Table 3) were much more highly correlated with median net cash income, from *all* sources, of rural farm families (.847\*\*), than with per worker net farm

income on all farms (.633\*\*).<sup>40</sup> Apparently the effect of the lesser response of farm-labor productivity on the Tennessee Valley's part-time farms was more than offset by greater non-farm income and by a greater response in the labor productivity of its commercial farms.<sup>41</sup>

#### V. SUMMARY AND CONCLUSIONS

The findings of our research at Vanderbilt on southern economic development and agriculture may be summarized as follows:

1. Within each area, today's more industrial counties had in 1850-60 an agriculture with certain significant advantages over the agriculture of today's less industrialized counties. Among these advantages were superior "original" agricultural land, certain differential windfall gains from new cash crops, and higher rates of capital formation in agriculture. Under conditions of perfect factor immobility, such "original" differences might have sufficed to account for today's wide income differentials within each study area. During the pre-industrial years 1860-1900, however, these earlier intercounty differences—for example, in agriculture's capital-labor ratio—virtually disappeared. Our analyses thus indicate that, despite imperfections in the factor markets, there were sufficient factor transfers (particularly labor) to eliminate these income differences attributable to differences in "original" physical endowment.

2. In 1900, today's more industrial counties did not have significantly different levels of farm capital per worker or farm output per worker from today's less industrial counties. By 1940, however,

<sup>40</sup> The corresponding coefficients for the Piedmont were .57\*\* and .58\*\*, respectively.

<sup>41</sup> Nicholls, "Industrial-Urban Development and Agricultural Adjustments . . .," *op. cit.*, pp. 147-48.

<sup>39</sup> Nicholls, "Industrial-Urban Development and Agricultural Adjustments . . .," *op. cit.*, pp. 145-46; and Tang, *op. cit.*, pp. 201-4.

today's more industrial counties had not only experienced differentially high rates of industrial-urban development but their agriculture also clearly enjoyed superior capital-labor ratios and higher farm-labor productivity. Although equilibrating factor transfers (as measured by relative rates of human migration) had meanwhile continued at substantial levels, these were obviously insufficient to prevent an increasing divergence in farm incomes and productivity between counties within each area. Instead, our analyses strongly support the view that local industrial-urban development made an important positive contribution to the efficiency of the local factor and product markets, thereby greatly facilitating the transfer of excess labor out of agriculture and of needed capital into agriculture within the immediate environs of the growing industrial center. Counties lacking such dynamic conditions of industrial-urban development were unable to hold their own through outmigration alone, even though such labor transfers had once sufficed to overcome income differences attributable only to "original" resource endowments.

3. The period 1940-54 permits the acid test for our conclusion that local areas cannot, in the absence of the industrial-urban development enjoyed by neighboring areas, overcome their differential disadvantage in farm incomes and productivity through outmigration alone. This most recent period provided not only sustained prosperity and full employment but conditions of war-induced resource mobility as well. The consequence was exceptionally high migration rates, particularly for our less indus-

trial counties. Yet, within each area, the more industrial counties have since 1940 been able to maintain most if not all of their relative advantage over the less advanced counties in per worker farm capital and farm output. Thus the ability of the more industrial counties to offer redundant farm workers a change of occupation without necessitating a change of residence was of paramount importance.

While our research at Vanderbilt has been applied to relatively small, homogeneous areas, I believe that its policy implications<sup>42</sup> have considerable relevance to the problems of much larger underdeveloped regions or countries (Brazil is a paramount example). First, industrial-urban development offers the major hope for solving the problem of low agricultural productivity, once prior problems of an inadequate food supply have been met. Second, insofar as it is not inconsistent with fundamental economies of location and scale, the more widely dispersed such industrial-urban development, the more generally can agricultural productivity be increased. Finally, particularly for these areas which lack the attributes required for sound industrialization, public policy must provide for facilitating farm-labor and farm-capital mobility at rates far in excess of those which can be expected under complete *laissez faire*.

<sup>42</sup> For a broader and more detailed translation of our research findings into policy recommendations for solving the low-income problems of the whole southern region of the United States, see my recent book, *Southern Tradition and Regional Progress* (Chapel Hill: University of North Carolina Press, 1960), pp. 164-76.



\_\_\_\_\_

12

13

14

15

---

## Part II The Role of Cities in National Development



---

## Introductory Note

Part I brought together materials on socioeconomic space and its transformations in the process of development. This space is not Perroux's "trivial" space of physical distance but the elastic space of human interactions where proximity and distance are functions of transportation and communication technologies, of culture, languages, and institutions, and of economic complementarities. Within this space, cities play a preeminent role as the nodal points in its structure, as the holders of power, and as the seedbeds of change. They are also seats of learning, engines of production, and confluences of power and control. In developed countries, which are predominantly urban societies, the cities relate primarily each to each; but developing countries are usually dual societies, where vast hinterlands are dominated by one or a few urban centers.

In view of their importance, we have devoted a special section to the role of cities in national development. Our central interest is the phenomenon of urbanization and the structuring of socioeconomic space by systems of cities rather than specific phenomena internal to cities. The six chapters that follow consequently respond primarily to questions on the role of cities in national development and attendant policies. \*

The brunt of several of these selections may be termed the confounding of accepted ideas about cities (that they become less efficient as they grow in size, that they lure migrants to conditions unfortunate to themselves and to others, that developing countries are overurbanized). Other common beliefs about urbanization, not represented in our selections for reasons of space, are also being challenged. They include the association of alienation with urban conditions, the morbidity of dense concentrations of population, and the reduction of agricultural production because of urban growth. Our selection of these pieces reflects a judgment on the quality of this research if not a conclusion about the validity of its findings and our opinion that most readers of this volume will already be familiar with the more commonly held beliefs and may find it interesting to have them challenged.

Robert Redfield and Milton Singer provide a sweeping historical perspective of "the part played by cities in the development, decline, and transformation of culture" (chapter 16). They are particularly concerned with the role of cities as moral, religious, and intellectual centers as distinct from their role as organizers of production and consumption, and wonder whether, in developing countries, cities can perform the former as well as the latter role.

Toshio Sanuki provides extraordinary documentation on the role of

information in the national system of cities based on Japanese data (chapter 17), following the stimulating original ideas of Richard L. Meier.<sup>1</sup> Although the details of Sanuki's analysis are not yet available, the regularity of the relations he displays is astounding. The implications he draws for policy are worth pondering, particularly as there is almost universal agreement that developed countries are now entering a new period, variously called a "service economy" or a "postindustrial society," based much more on information and social organization than on the mechanical aspects of material production. It may be observed that although there is almost universal agreement that this phenomenon is taking place, researchers seem to be tied hand and foot as to the urbanizing implications of this phenomenon and its policy correlates.

N. V. Sovani undertakes a critical exploration of the common idea that developing countries are "overurbanized" and concludes that there is no evidence for this assumption (chapter 18). In an analogous chapter, William Alonso examines the proposition that large metropolitan centers have grown beyond their level of efficiency, and he concludes that the evidence does not show this (chapter 19). Further, he expands the conception of the size of cities to include their proximity to other centers of population and concludes that "big and small must be qualified in their setting: whereas it may be quite good to be smaller in a dense setting, it may be quite necessary to be quite big in an isolated one. Policies for small and far, which are not uncommon, perhaps should be small and near, and big and far."

Joan Nelson looks at the political consequences of accelerated urban migration in newly industrializing countries (chapter 20). Does the apparent inability to absorb most of the incoming migrants productively forebode a period of threatening radicalization and political upheaval? A review of the literature fails to support this common contention. Nelson concludes, however, that the growing size of the urban proletariat and subproletariat may attract the interest of certain political groups in organizing and tapping the potential powers of the poor. The challenge, according to the author, is how governments will respond to the concrete and normally moderate demands of the dispossessed without sacrificing other development objectives.

John Friedmann and Flora Sullivan take a closer look at urban employment in developing countries (chapter 21). They propose a descriptive model of the labor market, as well as hypotheses concerning equilibrium unemployment, the pressure to subsistence in the urban economy, and the proletarianization of the urban labor force. A set of far-reaching proposals for policy conclude this chapter.

<sup>1</sup>Richard L. Meier, *A Communications Theory of Urban Growth* (Cambridge: MIT Press, 1962).

# 16 The Cultural Role of Cities

Robert Redfield and Milton B. Singer

This paper has as its purpose to set forth a framework of ideas that may prove useful in research on the part played by cities in the development, decline, or transformation of culture. "Culture" is used as in anthropology. The paper contains no report of research done. It offers a scheme of constructs; it does not describe observed conditions or processes; references to particular cities or civilizations are illustrative and tentative.

## Time Perspectives

The cultural role of cities may be considered from at least three different time perspectives. In the long-run perspective of human history as a single career,<sup>1</sup> the first appearance of cities marks a revolutionary change: the beginnings of civilization. Within this perspective cities remain the symbols and carriers of civilization wherever they appear. In fact the story of civilization may then be told as the story of cities — from those of the Ancient Near East through those of ancient Greece and Rome, medieval and modern Europe; and from Europe overseas to North and South America, Australia, the Far East, and back again to the modern Near East. In the short-run perspective we may study the cultural role of particular cities in relation to their local hinterlands of towns and villages.<sup>2</sup> The time span here is the several-year period of the field research or, at most, the lifespan of the particular cities that are studied. Between the long- and short-run perspectives, there is a middle-run perspective delimited by the life-history of the different civilizations within which cities have developed.<sup>3</sup> This is the perspective adopted when we consider the cultural bearings of urbanization within Mexican civilization,<sup>4</sup> or Chinese civ-

- (1) Robert Redfield, The Primitive World and its Transformations, Ithaca, New York, 1953, ix-xiii. W. N. Brown and others, "The Beginnings of Civilization," Journal of the American Oriental Society, Supplement No. 4, December, 1939, pp. 3-61.
- (2) Robert Redfield, The Folk Culture of Yucatan, Chicago: University of Chicago Press, 1941. This study, short-run in description, also aims to test some general ideas.  
Mandelbaum, David G. (ed.), "Integrated Social Science Research for India," Planning Memo., University of California, 1949.
- (3) Kroeber has recently discussed the problems of delimiting civilizations in his article, "The Delimitation of Civilizations," Journal of the History of Ideas, Vol. XIV (1953).  
Mark Jefferson, "Distribution of the world's city folk: a study in comparative civilization," Geographia, 1931.
- (4) Paul Kirchhoff, in "Four Hundred Years After: General Discussion of Acculturation, Social Change, and the Historical Provenience of Culture Elements," Heritage of Conquest by Sol Tax and others (Glencoe, Ill.: The Free Press, 1952), p. 254: "It seems to me that the fundamental characteristic of Mesoamerica was that it was a stratified society, one like ours or that of China, based on the axis of city and countryside. There was a native ruling class, with a class ideology and organization, which disappeared entirely; there were great cultural centers which, just as in our life, are so essential if you described the U. S. without New York, Chicago, etc., it would be absurd. The same thing happens

Reprinted from *Economic Development and Cultural Change*, vol. 3, no. 1 (October 1954), pp. 53-73, with permission of Professor Singer and The University of Chicago Press. © 1954 by The University of Chicago.

ilization or Indian civilization or Western civilization. It is a perspective usually of several thousand years and embraces within its orbit not just a particular city and its hinterland, but the whole pattern and sequence of urban development characteristic of a particular civilization and its cultural epochs.

While these three perspectives are clearly interrelated, research and analysis may concentrate primarily on one of them. Empirical ethnographic, sociological and geographical research on cities begins in the nature of the case with the short-run perspective, but the significance of such research increases as it becomes linked with ideas and hypotheses drawn from the other perspectives. One begins, say, with an empirical study of the origins, morphology, functions, and influence of an Asiatic city.<sup>5</sup> Then one may go on to look at this city as a link in the interaction of two distinct civilizations, and see the problem of urbanization in Asia generally as a problem in Westernization,<sup>6</sup> or the problem of Spanish-Indian acculturation of Mexico after the Conquest as a problem of de-urbanization and re-urbanization.<sup>7</sup> Finally, the canvas may be further enlarged to show both Western and Eastern cities as variants of a single and continuing cultural and historical process.<sup>8</sup> In this paper we propose to concentrate on the middle-run perspective, i.e., we shall analyze the role cities play in the formation, maintenance, spread, decline, and transformation of civilizations. We think that links with the long- and short-run perspectives will also emerge in the course of the analysis.

In the many useful studies of cities by urban geographers, sociologists, and ecologists we find frequent reference to "cultural functions" and "cultural centers."<sup>9</sup> Under these rubrics they generally include the religious, educational, artistic centers and activities, and distinguish them from administrative,

---

when you describe these centers in ancient Mexico. . . . It's not only the arts, crafts and sciences which constitute the great changes, but the basic form of the culture changing from a city structure to the most isolated form, which is, in my opinion, the most total and radical change anywhere in history. . . . When the city is cut off what is left over is attached as a subordinate to the new city-centered culture. . . ."

- (5) Ghosh, S., "The urban pattern of Calcutta," Economic Geography, 1950.  
Weulersse, J., "Antioche, un type de cité de l'Islam," Congr. int. de Géographie, Warsaw, 1934, III.  
D. R. Gadgil, Poona, A Socio-Economic Survey, Poona, 1945, 1952.
- (6) "Urbanization is part of the Europeanization that is spreading throughout the world," Mark Jefferson in reference (3) above. Kingsley Davis, The Population of India and Pakistan, Princeton, 1951, pp. 148-49; M. Zinkin, Asia and the West, London, 1951, Ch. 1, "Eastern Village and Western City."
- (7) Kirchhoff, op. cit.
- (8) See for this approach the books of V. Gordon Childe, and his article in Town Planning Review, XXI (1950) on "The Urban Revolution."
- (9) Grace M. Kneedler, "Functional types of cities," reprinted in Reader in Urban Sociology, edited by Paul K. Hatt and Albert J. Reiss, Jr., The Free Press, Glencoe, Illinois, 1951; R. E. Dickinson, The West European City, London: Routledge & Paul, 1951, pp. 253-54; Chauncey Harris, "A functional classification of cities in the United States," Geogr. Review, New York, 1943.

military, economic centers and functions. This usage of "cultural" is too narrow for the purpose of a comparative analysis of the role cities play in the transformations of the more or less integrated traditional life of a community. Economic and political centers and activities may obviously play as great a role in these processes as the narrowly "cultural" ones. Moreover, these different kinds of centers and activities are variously combined and separated and it is these varying patterns that are significant. In ancient civilizations the urban centers were usually political-religious or political-intellectual; in the modern world they are economic.<sup>10</sup> The mosque, the temple, the cathedral, the royal palace, the fortress, are the symbolic "centers" of the pre-industrial cities. The "central business district" has become symbolic of the modern urban center. In fact a cross-cultural history of cities might be written from the changing meanings of the words for city. "Civitas" in the Roman Empire meant an administrative or ecclesiastical district. Later, "city" was applied to the ecclesiastical center of a town — usually the cathedral. This usage still survives in names like "Île de la Cité" for one of the first centers of Paris. With the development of the "free cities," "city" came to mean the independent commercial towns with their own laws.<sup>11</sup> Today, "the city" of London is a financial center, and when Americans speak of "going to town" or "going downtown" they mean they are going to the "central business district." They usually think of any large city as a business and manufacturing center, whereas a Frenchman is more likely to regard his cities — certainly Paris — as "cultural centers."<sup>12</sup>

This symbolism is not of course a completely accurate designation of what goes on in the city for which it stands. The ecclesiastical centers were also in many cases centers of trade and of craftsmen, and the modern "central business district" is very apt to contain libraries, schools, art museums, government offices and churches, in addition to merchandising establishments and business offices. But allowing for this factual distortion, this symbolism does help us to separate two quite distinct cultural roles of cities, and provides a basis for classifying cities that is relevant to their cultural role. As a "central business district," the city is obviously a market-place, a place to buy and sell, "to do business" — to truck, barter and exchange with people who may be complete strangers and of different races, religions and creeds. The city here functions to work out largely impersonal relations among diverse cultural groups. As a religious or intellectual center, on the other hand, the city is a beacon for the faithful, a center for the learning, authority and perhaps doctrine that transforms the implicit "little traditions" of the local non-

(10) Gadgil, *The Industrial Revolution of India in Recent Times*, Oxford, 1944, pp. 6-12.

Spate and E. Ahmad, "Five cities of the Gangetic Plain. A cross-section of Indian cultural history," *Geog. Rev.*, 1950.

P. George, *La Ville*, Paris, 1952.

B. Rowland, *The Art and Architecture of India*, Penguin, Baltimore, 1953. Map showing ancient and historic art and religious centers, p. xvii.

Fei Hsiao-Tung, *China's Gentry, Essays in Rural-Urban Relations*, Chicago, University of Chicago Press, 1953, pp. 91-117.

(11) R. E. Dickinson, *op. cit.* (note 7), pp. 251-52; H. Pirenne, *Medieval Cities*.

(12) See article on "Urbanization" by W. M. Stewart in 14th edition of *Encyclopaedia Britannica* for some cultural variables in the definition of "city."



urban cultures into an explicit and systematic "great tradition." The varying cultural roles of cities, so separated and grouped into two contrasting kinds of roles with reference to the local traditions of the non-urban peoples, point to a distinction to which we shall soon return and to which we shall then give names.

#### Types of Cities

In the studies of economic historians (Pirenne, Dopsch) and in the studies of the currently significant factors for economic development (Hoselitz),<sup>13</sup> the functions of cities are considered as they effect change; but the change chiefly in view is economic change. Our attention now turns to the roles of cities in effecting change in the content and integration of ideas, interests and ideals.

The distinction Hoselitz takes from Pirenne between political-intellectual urban centers on the one hand and economic centers on the other, points in the direction of the distinction necessary to us in taking up the new topic. But the distinction we need does not fully emerge until we refine the classification by (1) separating the political function from the intellectual and (2) giving new content to the term "intellectual." Delhi, Quito and Peiping are to be contrasted, as Hoselitz says, with Bombay, Guayaquil and Shanghai because the former three cities are "political-intellectual centers" and the latter three are "economic centers." (The contrast of Rio to São Paulo is less clear.) Let us now add that there are cities with political functions and without significant intellectual functions: New Delhi (if it be fair to separate it from old Delhi), Washington, D. C. and Canberra (the new university there may require a qualification). Further, the intellectual functions of Delhi, Quito and Peiping (and Kyoto, Lhasa, Cuzco, Mecca, medieval Liège and Uaxactun) are to develop, carry forward, elaborate a long-established cultural tradition local to the community in which those cities stand. These are the cities of the literati: clerics, astronomers, theologians, imams and priests. New Delhi and Washington, D. C. do not have, significantly, literati; in spite of its schools and universities Washington is not a city of great intellectual leadership; these are cities without major intellectual functions. In respect to this lack, New Delhi and Washington, D. C., belong with cities with predominately economic functions. On the other hand, not a few old cities with economic functions have also the functions associated with the literati (Florence, medieval Timbuktoo; Thebes).

We have taken into consideration, in this expanded grouping, both cities of the modern era and cities of the time before the development of a world economy. It may be useful now to separate the two historic periods, retaining the distinction between cities of the literati, cities of entrepreneurs, and cities of the bureaucracy. The following grouping results:

#### BEFORE THE UNIVERSAL OEKUMENE (pre-industrial revolution, pre-Western expansion)

##### 1. Administrative-cultural cities (cities of the literati and the indigenous bureaucracy)

Peiping  
Lhasa  
Uaxactun  
Kyoto  
Liège  
Allahabad (?)

(13) B. Hoselitz, "The role of cities in the economic growth of underdeveloped countries," The Journal of Political Economy, vol. lxi (1953), esp. 198-99.

2. Cities of native commerce  
(cities of the entrepreneur)

Bruges  
Marseilles  
Lübeck  
Market towns of native West Africa  
Early Canton

AFTER THE UNIVERSAL OEKUMENE (post-industrial revolution,  
and post-Western expansion)

3. Metropolis-cities of the world-wide managerial and entrepreneurial class (Park's "cities of the main street of the world")

London  
New York  
Osaka  
Yokahama  
Shanghai  
Singapore  
Bombay

Lesser cities and towns, also carrying on the world's business, may be added here.

4. Cities of modern administration  
(cities of the new bureaucracies)

Washington, D. C.  
New Delhi  
Canberra

A thousand administrative towns, county seats, seats of British and French African colonial administration, etc.

What is the relationship of such a grouping to our topic: the role of cities in processes of cultural change?

The role of cities of Group 1 has already been stated. It is to carry forward, develop, elaborate a long-established local culture or civilization. These are cities that convert the folk culture into its civilized dimension.

But the cities of groups 2, 3, and 4 do not have, or do not have conspicuously and as their central effect, this role in the cultural process. They affect the cultural process in other ways. How? They are cities in which one or both of the following things are true: (1) the prevailing relationships of people and the prevailing common understandings have to do with the technical not the moral order,<sup>14</sup> with administrative regulation, business and technical convenience; (2) these cities are populated by people of diverse cultural origins removed from the indigenous seats of their cultures.

They are cities in which new states of mind, following from these characteristics, are developed and become prominent. The new states of mind are indifferent to or inconsistent with, or supersede or overcome, states of mind associated with local cultures and ancient civilizations. The intellectuals of these three groups of cities, if any, are intelligentsia rather than literati.<sup>15</sup>

(14) Robert Redfield, The Primitive World and Its Transformations, Ch. 3.

(15) Ibid, Ch. 3.

The distinction that is then basic to consideration of the cultural role of cities is the distinction between the carrying forward into systematic and reflective dimensions an old culture and the creating of original modes of thought that have authority beyond or in conflict with old cultures and civilizations. We might speak of the orthogenetic cultural role of cities as contrasted with the heterogenetic cultural role.

In both these roles the city is a place in which cultural change takes place. The roles differ as to the character of the change. Insofar as the city has an orthogenetic role, it is not to maintain culture as it was; the orthogenetic city is not static; it is the place where religious, philosophical and literary specialists reflect, synthesize and create out of the traditional material new arrangements and developments that are felt by the people to be outgrowths of the old. What is changed is a further statement of what was there before. Insofar as the city has a heterogenetic role, it is a place of conflict of differing traditions, a center of heresy, heterodoxy and dissent, of interruption and destruction of ancient tradition, of rootlessness and anomie. Cities are both these things, and the same events may appear to particular people or groups to be representative of what we here call orthogenesis or representative of heterogenesis. The predominating trend may be in one of the two directions, and so allow us to characterize the city, or that phase of the history of the city, as the one or the other. The lists just given suggest that the differences in the degree to which in the city orthogenesis or heterogenesis prevails are in cases strongly marked.

The presence of the market is not of itself a fact of heterogenetic change. Regulated by tradition, maintained by such customs and routines as develop over long periods of time, the market may flourish without heterogenetic change. In the medieval Muslim town we see an orthogenetic city; the market and the keeper of the market submitted economic activities to explicit cultural and religious definition of the norms. In Western Guatemala the people who come to market hardly communicate except with regard to buying and selling, and the market has little heterogenetic role. On the other hand the market in many instances provides occasions when men of diverse traditions may come to communicate and to differ; and also in the market occurs that exchange on the basis of universal standards of utility which is neutral to particular moral orders and in some sense hostile to all of them. The cities of Group 2, therefore, are cities unfavorable to orthogenetic change but not necessarily productive of heterogenetic change.

#### The City and the Folk Society<sup>16</sup>

The folk society may be conceived as that imagined combination of societal elements which would characterize a long-established, homogeneous, isolated and non-literate integral (self-contained) community; the folk culture is that society seen as a system of common understandings. Such a society can be approximately realized in a tribal band or village; it cannot be approximately realized in a city. What are characteristics of the city that may be conceived as a contrast to those of the folk society?

The city may be imagined as that community in which orthogenetic and heterogenetic transformations of the folk society have most fully occurred. The former has brought about the Great Tradition and its special intellectual class, administrative officers and rules closely derived from the moral and religious

(16) Robert Redfield, "The Natural History of the Folk Society," *Social Forces*, Vol. 31 (1953), pp. 224-28.

life of the local culture, and advanced economic institutions, also obedient to these local cultural controls. The heterogenetic transformations have accomplished the freeing of the intellectual, esthetic, economic and political life from the local moral norms, and have developed on the one hand an individualized expeditious motivation, and on the other a revolutionary, nativistic, humanistic or ecumenical viewpoint, now directed toward reform, progress and designed change.

As these two aspects of the effects of the city on culture may be in part incongruent with each other, and as in fact we know them to occur in different degrees and arrangements in particular cities, we may now review the classification of cities offered above so as to recognize at least two types of cities conceived from this point of view:

- A. The city of orthogenetic transformation: the city of the moral order; the city of culture carried forward. In the early civilizations the first cities were of this kind and usually combined this developmental cultural function with political power and administrative control. But it is to be emphasized that this combination occurred because the local moral and religious norms prevailed and found intellectual development in the literati and exercise of control of the community in the ruler and the laws. Some of these early cities combined these two "functions" with commerce and economic production; others had little of these. It is as cities of predominating orthogenetic civilization that we are to view Peiping, Lhasa, Uaxactun, fourteenth-century Liège.
- B. The city of heterogenetic transformation: the city of the technical order; the city where local cultures are disintegrated and new integrations of mind and society are developed of the kinds described above ("The heterogenetic role of cities"). In cities of this kind men are concerned with the market, with "rational" organization of production of goods, with expeditious relations between buyer and seller, ruler and ruled, and native and foreigner. In this kind of city the predominant social types are businessmen, administrators alien to those they administer, and rebels, reformers, planners and plotters of many varieties. It is in cities of this kind that priority comes to be given to economic growth and the expansion of power among the goods of life. The modern metropolis exhibits very much of this aspect of the city; the town built in the tropics by the United Fruit Company and the city built around the Russian uranium mine must have much that represents it; the towns of the colonial administration in Africa must show many of its features. Indeed, in one way or another, all the cities of groups 2, 3 and 4 (*supra*) are cities of the technical order, and are cities favorable to heterogenetic transformation of the moral order.<sup>17</sup>

This type of city may be subdivided into the administrative city, city of the bureaucracy (Washington, D. C., Canberra), and the city of the entrepreneur (Hamburg, Shanghai). Of course many cities exhibit both characteristics.

"In every tribal settlement there is civilization; in every city is the folk society." We may look at any city and see within it the folk society insofar as ethnic communities that make it up preserve folklike characteristics, and we may see in a town in ancient Mesopotamia or in aboriginal West Africa a half-

(17) In the heterogenetic transformation the city and its hinterland become mutually involved: the conservative or reactionary prophet in the country inveighs against the innovations or backslidings of the city; and the reformer with the radically progressive message moves back from Medina against Mecca, or enters Jerusalem.

way station between folk society and orthogenetic civilization. We may also see in every city its double urban characteristics: we may identify the institutions and mental habits there prevailing with the one or the other of the two lines of transformation of folk life which the city brings about. The heterogenetic transformations have grown with the course of history, and the development of modern industrial world-wide economy, together with the great movements of peoples and especially those incident to the expansion of the West, have increased and accelerated this aspect of urbanization. The later cities are predominantly cities of the technical order. We see almost side by side persisting cities of the moral order and those of the technical order: Peiping and Shanghai, Cuzco and Guayaquil, a native town in Nigeria and an administrative post and railway center hard by.

The ancient city, predominantly orthogenetic, was not (as remarked by W. Eberhard) in particular cases the simple outgrowth of a single pre-civilized culture, but was rather (as in the case of Loyang) a city in which conquered and conqueror lived together, the conqueror extending his tradition over the conquered, or accepting the latter's culture. What makes the orthogenetic aspect of a city is the integration and uniform interpretation of preceding culture, whether its origins be one or several. Salt Lake City and early Philadelphia, cities with much orthogenetic character, were established by purposive acts of founders. Salt Lake City created its own hinterland on the frontier (as pointed out by C. Harris). Other variations on the simple pattern of origin and development of a city from an established folk people can no doubt be adduced.

#### Transformation of Folk Societies: Primary Urbanization and Secondary Urbanization

The preceding account of different types of cities is perhaps satisfactory as a preliminary, but their cultural roles in the civilizations which they represent cannot be fully understood except in relation to the entire pattern of urbanization within that civilization, i.e., the number, size, composition, distribution, duration, sequence, morphology, function, rates of growth and decline, and the relation to the countryside and to each other of the cities within a civilization. Such information is rare for any civilization. In the present state of our knowledge it may be useful to guide further inquiry by assuming two hypothetical patterns of urbanization: primary and secondary.<sup>18</sup> In the primary phase a precivilized folk society is transformed by urbanization into a peasant society and correlated urban center. It is primary in the sense that the peoples making up the precivilized folk more or less share a common culture which remains the matrix too for the peasant and urban cultures which develop from it in the course of urbanization. Such a development, occurring slowly in communities not radically disturbed, tends to produce a "sacred culture" which is gradually transmuted by the *literati* of the cities into a "Great Tradition." Primary urbanization thus takes place almost entirely within the framework of a core culture that develops, as the local cultures become urbanized and transformed, into an indigenous civilization. This core culture dominates the civilization despite occasional intrusions of foreign peoples and cultures. When the encounter with other peoples and civilizations is too rapid and intense an indigenous civilization may be destroyed by de-urbanization or be

(18) This distinction is an extension of the distinction between the primary and secondary phases of folk transformations in Redfield, The Primitive World and Its Transformations, p. 41.

variously mixed with other civilizations.<sup>19</sup>

This leads to the secondary pattern of urbanization: the case in which a folk society, precivilized, peasant, or partly urbanized, is further urbanized by contact with peoples of widely different cultures from that of its own members. This comes about through expansion of a local culture, now partly urbanized, to regions inhabited by peoples of different cultures, or by the invasion of a culture-civilization by alien colonists or conquerors. This secondary pattern produces not only a new form of urban life in some part in conflict with local folk cultures but also new social types in both city and country. In the city appear "marginal" and "cosmopolitan" men and an "intelligentsia"; in the country various types of marginal folk: enclaved-, minority-, imperialized-, transplanted-, remade-, quasi-folk, etc., depending on the kind of relation to the urban center.

This discussion takes up a story of the contact of peoples at the appearance of cities. But, here parenthetically, it is necessary to note that even before the appearance of cities the relations between small and primitive communities may be seen as on the one hand characterized by common culture and on the other by mutual usefulness with awareness of cultural difference. The "primary phase of urbanization" is a continuation of the extension of common culture from a small primitive settlement to a town and its hinterland, as no doubt could be shown for parts of West Africa. The "secondary phase of urbanization" is begun, before cities, in the institutions of travel and trade among local communities with different cultures. In Western Guatemala today simple Indian villagers live also in a wider trade-community of pluralistic cultures;<sup>20</sup> we do not know to what extent either the pre-Columbian semi-urban centers or the cities of the Spanish-modern conquerors and rulers, have shaped this social system; it may be that these people were already on the way to secondary urbanization before any native religious and political center rose to prominence.

While we do not know universal sequences within primary or secondary urbanization, it is likely that the degree to which any civilization is characterized by patterns of primary or secondary urbanization depends on the rate of technical development and the scope and intensity of contact with other cultures. If technical development is slow and the civilization is relatively isolated, we may expect to find a pattern of primary urbanization prevailing. If, on the other hand, technical development is rapid and contacts multiple and intense, secondary urbanization will prevail.

It may be that in the history of every civilization there is, of necessity, secondary urbanization. In modern Western civilization conditions are such as to make secondary urbanization the rule. But even in older civilizations it is not easy to find clear-cut examples of primary urbanization — because of multiple interactions, violent fluctuations in economic and military fortunes, conflicts and competition among cities and dynasties, and the raids of nomads. The Maya before the Spanish Conquest are perhaps a good example of primary urbanization.<sup>21</sup> The cases of the Roman, Greek, Hindu, Egyptian and Mesopo-

(19) Kirchhoff, *op. cit.*

(20) R. Redfield, "Primitive Merchants of Guatemala," *Quarterly Journal of Inter-American Relations*, Vol. 1, No. 4, 1939, pp. 48-49.

(21) Redfield, *The Primitive World and Its Transformations*, pp. 58-73. See also Morley, *The Ancient Maya*, and Thomas Gann and J. Eric Thompson, *The History of the Maya*, New York, 1931.

tamian civilizations, although characterized by distinctive indigenous civilizations, are nevertheless complex because little is known about the degree of cultural homogeneity of the peoples who formed the core cultures and because as these civilizations became imperial they sought to assimilate more and more diverse peoples. Alternatively the irritant "seed" of a city may have been sown in some of them by the conquering raid of an outside empire, the desire to copy another empire in having a capital, or simple theft from another people — with the subsequent development around this seed of the "pearl" of a relatively indigenous, primary urban growth, sending out its own imperial secondary strands in due time. Thus while Rome, Athens, Chang-An and Loyang in early China and Peiping in later, Pataliputra and Benares, Memphis and Thebes, Nippur and Ur, may have been for a time at least symbolic vehicles for loyalty to the respective empires and indigenous civilizations, it was not these relatively "orthogenetic" cities but the mixed cities on the periphery of an empire — the "colonial cities" which carried the core culture to other peoples. And in such cities, usually quite mixed in character, the imperial great tradition was not only bound to be very dilute but would also have to meet the challenge of conflicting local traditions. At the imperial peripheries, primary urbanization turns into secondary urbanization.<sup>22</sup>

Similar trends can be perceived in modern times: Russian cities in Southern Europe and Asia appear to be very mixed,<sup>23</sup> non-Arabic Muslim cities have developed in Africa and South Asia, and the colonial cities of the European powers admit native employees daily at the doors of their skyscraper banks. Possibly the nuclear cultures are homogeneous and create indigenous civilizations but as they expand into new areas far afield from the home cultures they have no choice but to build "heterogenetic" cities.

Modern "colonial" cities (e.g., Jakarta, Manila, Saigon, Bangkok, Singapore, Calcutta) raise the interesting question whether they can reverse from the "heterogenetic" to the "orthogenetic" role. For the last one hundred or more years they have developed as the outposts of imperial civilizations, but as the countries in which they are located achieve political independence, will the cities change their cultural roles and contribute more to the formation of a civilization indigenous to their areas? Many obstacles lie in the path of such a course. These cities have large, culturally diverse populations, not necessarily European, for example, the Chinese in Southeast Asia, Muslims and Hindu refugees from faraway provinces, in India; they often have segregated ethnic quarters, and their established administrative, military and economic functions are not easily changed. Many new problems have been created by a sudden influx of postwar refugee populations, and the cities' changing positions in national and global political and economic systems. While many of these colonial cities have been centers of nationalism and of movements for revival

- 
- (22) The case of China is particularly striking, since the evidence for a dominant core culture is unmistakable but its relation to local cultures which may have been its basis is unknown. See Chi Li, *The Formation of the Chinese People*, Cambridge, Harvard University Press, 1928, and Wolfram Eberhard, *Early Chinese Cultures and their Development*, Smithsonian Institution Annual Report, 1937, Washington, 1938.

For a good study of imperial "spread" and "dilution," see A. H. M. Jones, *The Greek City from Alexander to Justinian*, Oxford, 1940.

- (23) Chauncy Harris, "Ethnic groups in cities of the Soviet Union," *Geog. Rev.*, 1945.



of the local cultures, they are not likely to live down their "heterogenetic" past.<sup>24</sup>

The Cultural Consequences of Primary and  
Secondary Urbanization

The discussion of primary and secondary urbanization above has been a bare outline. It may be filled in by reference to some postulated consequences of each type of process. The most important cultural consequence of primary urbanization is the transformation of the Little Tradition into a Great Tradition. Embodied in "sacred books" or "classics," sanctified by a cult, expressed in monuments, sculpture, painting, and architecture, served by the other arts and sciences, the Great Tradition becomes the core culture of an indigenous civilization and a source, consciously examined, for defining its moral, legal, aesthetic and other cultural norms. A Great Tradition describes a way of life and as such is a vehicle and standard for those who share it to identify with one another as members of a common civilization. In terms of social structure, a significant event is the appearance of literati, those who represent the Great Tradition. The new forms of thought that now appear and extend themselves include reflective and systematic thought; the definition of fixed idea-systems (theologies, legal codes); the development of esoteric or otherwise generally inaccessible intellectual products carried forward, now in part separate from the tradition of the folk; and the creation of intellectual and aesthetic forms that are both traditional and original (cities of the Italian Renaissance; development of "rococo" Maya sculpture in the later cities).

In government and administration the orthogenesis of urban civilization is represented by chiefs, rulers and laws that express and are closely controlled by the norms of the local culture. The chief of the Crow Indians, in a pre-civilized society, and the early kings of Egypt, were of this type. The Chinese emperor was in part orthogenetically controlled by the Confucian teaching and ethic; in some part he represented a heterogenetic development. The Roman pro-consul and the Indian Service of the United States, especially in certain phases, were more heterogenetic political developments.

Economic institutions of local cultures and civilizations may be seen to be orthogenetic insofar as the allocation of resources to production and distribution for consumption are determined by the traditional system of status and by the traditional specific local moral norms. The chief's yam house in the Trobriands is an accumulation of capital determined by these cultural factors. In old China the distribution of earnings and "squeeze" were distributed according to familial obligations: these are orthogenetic economic institutions and practices. The market, freed from controls of tradition, status and moral rule, becomes the world-wide heterogenetic economic institution.

In short, the trend of primary urbanization is to co-ordinate political, economic, educational, intellectual and aesthetic activity to the norms provided by the Great Traditions.

The general consequence of secondary urbanization is the weakening or supersession of the local and traditional cultures by states of mind that are

(24) D. W. Fryer, "The 'million city' in Southeast Asia," *Geog. Rev.*, Oct., 1953; J. E. Spencer, "Changing Asiatic cities," *Geog. Rev.*, Vol. 41 (1951). This last is a summary of an article by Jean Chesneaux. See also *Record of the XXVIIth Meeting of the International Institute of Differing Civilizations*, Brussels, 1952, esp. papers by R. W. Steel and K. Neys.



incongruent with those local cultures. Among these are to be recognized:

1. The rise of a consensus appropriate to the technical order: i.e., based on self-interest and pecuniary calculation, or on recognition of obedience to common impersonal controls, characteristically supported by sanctions of force. (This in contrast to a consensus based on common religious and non-expediential moral norms.) There is also an autonomous development of norms and standards for the arts, crafts, and sciences.

2. The appearance of new sentiments of common cause attached to groups drawn from culturally heterogeneous backgrounds. In the city proletariats are formed and class or ethnic consciousness is developed, and also new professional and territorial groups. The city is the place where ecumenical religious reform is preached (though it is not originated there). It is the place where nationalism flourishes. On the side of social structure, the city is the place where new and larger groups are formed that are bound by few and powerful common interests and sentiments in place of the complexly inter-related roles and statuses that characterize the groups of local, long-established culture. Among social types that appear in this aspect of the cultural process in the city are the reformer, the agitator, the nativistic or nationalistic leader, the tyrant and his assassin, the missionary and the imported school teacher.

3. The instability of viewpoint as to the future, and emphasis on prospective rather than retrospective view of man in the universe. In cities of predominantly orthogenetic influence, people look to a future that will repeat the past (either by continuing it or by bringing it around again to its place in the cycle). In cities of predominantly heterogenetic cultural influence there is a disposition to see the future as different from the past. It is this aspect of the city that gives rise to reform movements, forward-looking myths, and planning, revolutionary or melioristic. The forward-looking may be optimistic and radically reformistic; it may be pessimistic, escapist, defeatist or apocalyptic. In the city there are Utopias and counter-Utopias. Insofar as these new states of mind are secular, worldly, they stimulate new political and social aspiration and give rise to policy.

#### Consequences for World View, Ethos, and Typical Personality

The difference in the general cultural consequences of primary and secondary urbanization patterns may be summarily characterized by saying that in primary urbanization, all phases of the technical order (material technology, economy, government, arts, crafts, and sciences) are referred, in theory at least, to the standards and purposes of a moral order delineated in the Great Tradition, whereas in secondary urbanization different phases of the technical order are freed from this reference and undergo accelerated autonomous developments. With respect to this development, the moral order, or rather orders, for there are now many competing ones, appears to lag.<sup>25</sup>

There is another way of describing these differences: in terms of the consequences of the two kinds of urbanization for changes in world view, ethos, and typical personality.<sup>26</sup> To describe the consequences in these terms is to describe them in their bearings and meanings for the majority of individual

(25) Redfield, *The Primitive World and Its Transformations*, pp. 72-83.

(26) For a further discussion of these concepts, see Redfield, *ibid.*, Ch. 4, and Redfield, *The Little Community*, University of Chicago Press (forthcoming), Chs. 5 and 6 on personality and mental outlook.

selves constituting the society undergoing urbanization. We now ask, how do primary and secondary urbanization affect mental outlook, values and attitudes, and personality traits? These are in part psychological questions, for they direct our attention to the psychological aspects of broad cultural processes.

There are many accounts of the psychological consequences of urbanization. These have described the urban outlook, ethos, and personality as depersonalized, individualized, emotionally shallow and atomized, unstable, secularized, blasé, rationalistic, cosmopolitan, highly differentiated, self-critical, time-coordinated, subject to sudden shifts in mood and fashion, "other-directed," etc.<sup>27</sup> The consensus in these descriptions and their general acceptance by social scientists seem great enough to indicate that there probably is a general psychological consequence of urbanization, although it cannot be precisely described and proven. We should, however, like to suggest that the "urban way of life" that is described in the characterizations to which we refer is primarily a consequence of secondary urbanization and of that in a particular critical stage when personal and cultural disorganization are greatest. To see these consequences in perspective, it is necessary to relate them on the one hand to the consequences of primary urbanization and on the other to those situations of secondary urbanization that produce new forms of personal and cultural integration. Most of all it is necessary to trace the continuities as well as the discontinuities in outlook, values, and personality, as we trace the transformation of folk societies into their civilized dimension. The "peasant" is a type that represents an adjustment between the values of the precivilized tribe and those of the urbanite. The "literati" who fashion a Great Tradition do not repudiate the values and outlook of their rural hinterland but systematize and elaborate them under technical specialization. The cosmopolitan "intelligentsia" and "sophists" of the metropolitan centers have a prototype in the "heretic" of the indigenous civilization. And even the most sophisticated urban centers are not without spiritualists, astrologers and other practitioners with links to a folk-like past.<sup>28</sup>

The connections between the folk culture, the Great Tradition, and the sophisticated culture of the heterogenetic urban centers can be traced not only in the continuities of the historical sequence of a particular group of local cultures becoming urbanized and de-urbanized, but they also can be traced in the development of two distinct forms of cultural consciousness which appear in these transformations.

(27) See L. Wirth, "Urbanism as a way of life," and G. Simmel, "The metropolis and mental life," both reprinted in Hatt and Reiss, Reader in Urban Sociology; E. Fromm, Escape from Freedom, David Riesman and collaborators, The Lonely Crowd, and A. Kroeber, Anthropology, 1948, sec. 121. For the effects of urban life on time-coordination, see H. A. Hawley, Human Ecology, Ch. 15, and P. Hallowell, "Temporal orientations in western and non-western cultures," (?), American Anthropologist, Vol. 39, 1937.

(28) Redfield, The Folk Culture of Yucatan, Ch. 11; R. E. Park, "Magic, Mentality, and City Life," reprinted in Park, Human Communities.

N. C. Chaudhuri, The Autobiography of an Unknown Indian, Macmillan, 1951, gives some interesting observations on the survival of "folk" beliefs and practices among the people of Calcutta, pp. 361-62.

P. Maasson-Oursel, "La Sophistique. Etude de philosophie comparée," Revue de métaphysique et de morale, 23 (1916), pp. 343-62.

Cultural Integration Between City and Country

From what has been said about primary and secondary urbanization it follows that city and country are more closely integrated, culturally, in the primary phase of urbanization than in the secondary phase. Where the city has grown out of a local culture, the country people see its ways as in some important part a form of their own, and they feel friendlier toward the city than do country people ruled by a proconsul from afar. The stereotype of "the wicked city" will be stronger in the hinterlands of the heterogenetic cities than in those of the orthogenetic cities. Many of these are sacred centers of faith, learning, justice and law.

Nevertheless, even in primary urbanization a cultural gap tends to grow between city and country. The very formation of the Great Tradition introduces such a gap. The literati of the city develop the values and world view of the local culture to a degree of generalization, abstraction and complexity incomprehensible to the ordinary villager, and in doing so leave out much of the concrete local detail of geography and village activity. The Maya Indian who lived in some rural settlement near Uaxactun could not have understood the calendrical intricacies worked out in that shrine-city by the priests; and the rituals performed at the city-shrine had one high level of meaning for the priest and another lower meaning, connecting with village life at some points only, for the ordinary Indian.

On the other hand, primary urbanization involves the development of characteristic institutions and societal features that hold together, in a certain important measure of common understanding, the Little Tradition and the Great Tradition. We may refer to the development of these institutions and societal features as the universalization of cultural consciousness — meaning by "universalization," the preservation and extension of common understanding as to the meaning and purpose of life, and sense of belonging together, to all the people, rural or urban, of the larger community. Some of the ways in which this universalization takes place are suggested in the following paragraphs. The examples are taken chiefly from India; they probably have considerable cross-cultural validity.

1. The embodiment of the Great Tradition in "sacred books" and secondarily in sacred monuments, art, icons, etc. Such "sacred scriptures" may be in a language not widely read or understood; nevertheless they may become a fixed point for the worship and ritual of ordinary people. The place of the "Torah" in the lives of Orthodox Jews, the Vedas among orthodox Hindus, the "Three Baskets" for Buddhists, the thirteen classics for Confucianists, the Koran for Muslims, the stelae and temples of the ancient Maya, are all examples of such sacred scriptures, although they may vary in degree of sacredness and in canonical status.

2. The development of a special class of "literati" (priests, rabbis, Imams, Brahmins) who have the authority to read, interpret, and comment on the sacred scriptures. Thus the village Brahmin who reads the Gita for villagers at ceremonies mediates a part of the Great Tradition of Hinduism for them.

The mediation of a great tradition is not always this direct. At the village level it may be carried in a multitude of ways — by the stories parents and grandparents tell children, by professional reciters and storytellers, by dramatic performances and dances, in songs and proverbs, etc.

In India the epics and puranas have been translated into the major

regional languages and have been assimilated to the local cultures. This interaction of a "great tradition" and the "little tradition" of local and regional cultures needs further study, especially in terms of the professional and semi-professional "mediators" of the process.

3. The role of leading personalities who because they themselves embody or know some aspects of a Great Tradition succeed through their personal position as leaders in mediating a Great Tradition to the masses of people. There is a vivid account of this process in Jawarhalal Nehru's Discovery of India, in which he describes first how he "discovered" the Great Tradition of Indian in the ruins of Mohenjo-Daro and other archeological monuments, her sacred rivers and holy cities, her literature, philosophy, and history. And then he describes how he discovered the "little traditions" of the people and the villages, and how through his speeches he conveyed to them a vision of Bharat Mata — Mother India — that transcended the little patches of village land, people, and customs.<sup>29</sup>

4. Nehru's account suggests that actual physical places, buildings and monuments — especially as they become places of sacred or patriotic pilgrimage — are important means to a more universalized cultural consciousness and the spread of a Great Tradition. In India this has been and still is an especially important universalizing force. The sanctity of rivers and the purifying powers of water go all the way back to the Rig Veda. The Buddhists — who may have started the practice of holy pilgrimages — believed that there were four places that the believing man should visit with awe and reverence: Buddha's birth place, the site where he attained illumination or perfect insight, the place where the mad elephant attacked him, and the place where Buddha died. In the Mahabharata, there is a whole book on the subject of holy places (Arareyaka Book). Even a sinner who is purified by holy water will go to heaven. And the soul ready for moksha will surely achieve it if the pilgrim dies on a pilgrimage.<sup>30</sup> Today the millions of pilgrims who flock to such preeminent holy spots as Allahabad or Banaras create problems of public safety and urban over-crowding, but they, like Nehru, are also discovering the Bharat Mata beyond their villages.

In India "sacred geography" has also played an important part in determining the location and layout of villages and cities and in this way has created a cultural continuity between countryside and urban centers. In ancient India, at least, every village and every city had a "sacred center" with temple, tank, and garden. And the trees and plants associated with the sacred shrine were also planted in private gardens, for the households too had their sacred center; the house is the "body" of a spirit (Varta Purusha) just as the human body is the "house" of the soul.<sup>31</sup>

At each of these levels — of household, village, and city — the "sacred

(29) Jawarhalal Nehru, The Discovery of India, John Day, New York, 1946, pp. 37-40, 45-51.

(30) D. Patil, Cultural History from the Vāya Purāna, Poona, 1946, Appendix B.

(31) C. P. V. Ayyar, Town Planning in the Ancient Dekkan, Madras, no date, with an introduction by Patrick Geddes. See also Patrick Geddes in India, ed. J. Tyrwhitt, London, 1947.

N. V. Ramanayya, An Essay on the Origin of the South Indian Temple, Madras, 1930, and Stella Kramrisch, The Hindu Temple, Calcutta, 1946.

center" provides the forum, the vehicle, and the content for the formation of distinct cultural identities — of families, village, and city. But as individuals pass outward, although their contacts with others become less intimate and less frequent, they nevertheless are carried along by the continuity of the "sacred centers," feeling a consciousness of a single cultural universe where people hold the same things sacred, and where the similarities of civic obligations in village and city to maintain tanks, build public squares, plant fruit trees, erect platforms and shrines, is concrete testimony to common standards of virtue and responsibility.

Surely such things as these — a "sacred scripture," and a sacred class to interpret it, leading personalities, "sacred geography" and the associated rites and ceremonies — must in any civilization be important vehicles for the formation of that common cultural consciousness from which a Great Tradition is fashioned and to which it must appeal if it is to stay alive. It is in this sense that the universalization of cultural consciousness is a necessary ingredient in its formation and maintenance. Moreover, as the discussion of the role of "sacred geography" in the formation of Hinduism has intimated, this process does not begin only at the point where the villager and the urbanite merge their distinct cultural identities in a higher identity, but is already at work at the simpler levels of family, caste and village, and must play an important part in the formation and maintenance of the Little Tradition at these levels.<sup>32</sup>

---

H. Rao, "Rural habitation in South India," Quarterly Journal of the Mythic Society, 14.

J. M. Linton Bogle, Town Planning in India, Oxford University Press, 1929.

Mudgett and others, Banaras: Outline of a Master Plan, prepared by Town and Village Planning Office, Lucknow.

- (32) See Robert Redfield, The Little Community, (ms. to be published 1954) Ch. 8, on the little community "As a community within communities."

In addition to the above factors, it has been usual to single out special items of content of the world view and values of a Great Tradition as explanations of the "Universalization" of Great Traditions. It has been frequently argued, e.g., that religions which are monotheistic and sanction an "open class" social system will appeal more to ordinary people and spread faster than those which are polytheistic and which sanction "caste" systems. (See e.g., H. J. Kissling, "The sociological and educational role of the Dervish orders in the Ottoman Empire," in G. von Grunebaum (ed.), Studies in Islamic Cultural History.) F. S. C. Northrop and Arnold Toynbee both attach great importance to the ideological content of cultures as factors in their spread, although they come out with different results. It may be that such special features of content are important in the formation and spread of some particular religions at some particular time, but it is doubtful that they would have the same role in different civilizations under all circumstances. In his recent study of the Coorgs of South India, Srinivas argues with considerable plausibility that the spread of Hinduism on an all-India basis has depended on its polytheism, which has made it easy to incorporate all sorts of alien deities, and on a caste system which assimilates every new cultural or ethnic group as a special caste.<sup>28</sup>

Another difficulty about using special features of content of some partic-

The integration of city and country in the secondary phase of urbanization cannot rest on a basic common cultural consciousness or a common culture, for there is none. Rural-urban integration in this phase of urbanization rests primarily on the mutuality of interests and on the "symbiotic" relations that have often been described.<sup>33</sup> The city is a "service station" and amusement center for the country, and the country is a "food basket" for the city. But while the diversity of cultural groups and the absence of a common culture makes the basis of the integration primarily technical, even this kind of integration requires a kind of cultural consciousness to keep it going. We refer to the consciousness of cultural differences and the feeling that certain forms of inter-cultural association are of great enough benefit to override the repugnance of dealing with "foreigners." We may call this an "enlargement of cultural horizons sufficient to become aware of other cultures and of the possibility that one's own society may in some ways require their presence. To paraphrase Adam Smith, it is not to the interest of the (Jewish) baker, the (Turkish) carpet-dealer, the (French) hand laundry, that the American Christian customer looks when he patronizes them, but to his own.

This is the practical psychological basis for admission of the stranger and tolerance of foreign minorities, even at the level of the folk society.<sup>34</sup> In a quotation from the *Institutions of Athens*, which Toynbee has, perhaps ironically, titled "Liberté-Egalité-Fraternité," we are told that the reason why Athens has "extended the benefits of Democracy to the relations between slaves and freemen and between aliens and citizens" is that "the country requires permanent residence of aliens in her midst on account both of the multiplicity of trades and of her maritime activities."<sup>35</sup>

When all or many classes of a population are culturally strange to each other and where some of the city populations are culturally alien to the country populations, the necessity for an enlarged cultural consciousness is obvious. In societies where social change is slow, and there has developed an adjustment of mutual usefulness and peaceful residence side by side of groups culturally different but not too different, the culturally complex society may be relatively stable.<sup>36</sup> But where urban development is great, such conditions

---

ular tradition as a general explanation of the formation and maintenance of any Great Tradition is that one inevitably selects features that have been crystallized only after a long period of historical development and struggle. These are more relevant as factors in explaining further development and spread than they are in explaining the cultural-psychological processes that have accompanied primary urbanization. The "universalization" of universal faiths takes us into the realm of secondary urbanization where diverse and conflicting cultures must be accommodated.

- (33) R. E. Park, "Symbiosis and socialization: a frame of reference for the study of society," reprinted in *Human Communities*, Free Press, Glencoe, 1952.
- (34) Robert Redfield, *The Primitive World and Its Transformations*, pp. 33-34, for the institutionalization of hospitality to strangers in peasant societies.
- (35) Arnold Toynbee, *Greek Civilization and Character*, Beacon Press, Boston, 1950, pp. 48-49. See also David G. Mandelbaum, "The Jewish way of life in Cochín," *Jewish Social Studies*, Vol. I (1939).
- (36) Redfield, "Primitive Merchants of Guatemala."

are apt to be unstable. Each group may be perpetually affronted by the beliefs and practices of the other groups. Double standards of morality will prevail, since each cultural group will have one code for its "own kind" and another for the "outsiders." This simultaneous facing both inward and outward puts a strain on both codes. There may then be present the drives to proselytize, to withdraw and dig in, to persecute and to make scapegoats; there may even be fear of riot and massacre. In such circumstances the intellectuals become the chief exponents of a "cosmopolitan" enlarged cultural consciousness, inventing formulas of universal toleration and the benefits of mutual understanding, and extolling the freedom to experiment in different ways of life. But they do not always prevail against the more violent and unconvinced crusaders for some brand of cultural purity.

In primary urbanization when technical development was quite backward, a common cultural consciousness did get formed. The travelling student, teacher, saint, pilgrim or even humble villager who goes to the next town may be startled by strange and wonderful sights, but throughout his journey he is protected by the compass of the common culture from cultural shock and disorientation. In ancient times students and teachers came from all over India and even from distant countries to study at Taxila, just as they came from all over Greece to Athens. In secondary urbanization, especially under modern conditions, technical developments in transportation, travel and communication enormously facilitate and accelerate cultural contacts. The effects of this on common cultural consciousness are not easy briefly to characterize. They make the more traditional cultural differences less important. They provide a wide basis of common understanding with regard to the instruments and practical means of living. It is at least clear that the integration of country and city that results is not the same kind of sense of common purpose in life that was provided to rural-urban peoples through the institutions mediating Little and Great Traditions referred to above. At this point the enquiry approaches the questions currently asked about the "mass culture" of modern great societies.

Cities as Centers of Cultural Innovation,  
Diffusion, and Progress

It is a commonly stated view that the city rather than the country is the source of cultural innovations, that such innovations diffuse outward from city to country, and that the "spread" is more or less inverse to distance from the urban center.<sup>37</sup> The objection to this view is not that it is wrong — for there

- (37) P. Sorokin and C. Zimmerman, *Principles of Rural-Urban Sociology*, New York, Henry Holt and Co., 1929, Ch. 17, "The role of the city and the country in innovation, disruption, and preservation of the national culture."

Chabot, G., "Les zones d'influence d'une ville," *Congr. int. de Geog.*, Paris, 1931, III, pp. 432-37.

Jefferson, Mark, "The law of the primate city," *Geog. Rev.*, 1939, 226-32.

Spate, O. H. K., "Factors in the development of capital cities," *Geog. Rev.*, 1942, pp. 622-31.

R. E. Park, "The urban community as a spatial pattern and a moral order," "Newspaper circulation and metropolitan regions," both reprinted in Park, *Human Communities*.



is much evidence that would seem to support it — but that the limits and conditions of its validity need to be specified. It seems to assume for example that in the processes of cultural change, innovation, and diffusion, "city" and "country" are fixed points of reference which do not have histories, or interact, and are not essentially related to larger contexts of cultural change. Yet such assumptions — if ever true — would hold only under the most exceptional and short-run conditions. It is one thing to say that a large metropolitan city is a "center" of cultural innovation and diffusion for its immediate hinterland at a particular time; it is another to ask how that center itself was formed, over how long a period and from what stimuli. In other words, as we enlarge the time span, include the rise and fall of complex distributions of cities, allow for the mutual interactions between them and their hinterlands, and also take account of interactions with other civilizations and their rural-urban patterns, we find that the processes of cultural innovation and "flow" are far too complex to be handled by simple mechanical laws concerning the direction, rate, and "flow" of cultural diffusion between "city" and "country." The cities themselves are creatures as well as creators of this process, and it takes a broad cross-cultural perspective to begin to see what its nature is. While this perspective may not yield simple generalizations about direction and rates of cultural diffusion to widen the viewpoint as here suggested may throw some light on the processes of cultural change, including the formation and cultural "influence" of cities.

In a primary phase of urbanization, when cities are developing from folk societies, it seems meaningless to assert, e.g., that the direction of cultural flow is from city to country. Under these conditions a folk culture is transformed into an urban culture which is a specialization of it, and if we wish to speak of "direction of flow" it would make more sense to see the process as one of a series of concentrations and nucleations within a common field. And as these concentrations occur, the common "Little Tradition" has not become inert; in fact, it may retain a greater vitality and disposition to change than the systematized Great Tradition that gets "located" in special classes and in urban centers. From this point of view the spatial and mechanical concepts of "direction" and "rate" of flow, etc., are just metaphors of the processes involved in the formation of a Great Tradition. The cultural relations between city and country have to be traced in other terms, in terms of socio-cultural history and of cultural-psychological processes. Physical space and time may be important obstacles and facilitators to these processes but they are not the fundamental determinants of cultural "motion" as they are of physical motion.

Under conditions of secondary urbanization, the spatial and mechanical concepts seem more appropriate because people and goods are more mobile and the technical development of the channels of transportation and communication is such as to permit highly precise measurement of their distributions and of "flows." But here too we may be measuring only some physical facts whose cultural significance remains indeterminate, or, at most, we may be documenting only a particularly recent cultural tendency to analyze intercultural relations in quantitative, abstract, and non-cultural terms. The assumption of a continuous and quantitatively divisible "diffusion" from a fixed urban center is unrealistic.

We may see Canton or Calcutta as a center for the diffusion of Western culture into the "East." We may also see these cities as relatively recent

---

Hiller, "Extension of urban characteristics into rural areas," *Rural Sociology*, Vol. 6 (1941).



metropolitan growths, beginning as minor outliers of Oriental civilizations and then attracting both foreign and also uprooted native peoples, varying in fortune with world-wide events, and becoming at last not so much a center for the introduction of Western ways as a center for nativistic and independence movements to get rid of Western control and dominance. "Everything new happens at Canton," is said in China. We have in such a case not simple diffusion, or spread of urban influence from a city, but rather a cultural interaction which takes place against a background of ancient civilization with its own complex and changing pattern of urbanization now coming into contact with a newer and different civilization and giving rise to results that conform to neither.

The city may be regarded, but only very incompletely, as a center from which spreads outward the idea of progress. It is true that progress, like the ideologies of nationalism, socialism, communism, capitalism and democracy, tends to form in cities and it is in cities that the prophets and leaders of these doctrines are formed. Yet the states of mind of Oriental and African peoples are not copies of the minds of Western exponents of progress or of one or another political or economic doctrine. There is something like a revolution of mood and aspiration in the non-European peoples today.<sup>38</sup> The Easterner revolts against the West; he does not just take what can be borrowed from a city; he does sometimes the opposite: the Dutch language is set aside in Indonesia; there, anthropology, because associated with Dutch rule, does not spread from any city but is looked on with suspicion as associated with Dutch rule. Moreover, the influence of the West does not simply move outward from cities; it leap-frogs into country regions; a city reformer in Yucatan, Carrillo Puerto, arouses village Indians to join his civil war for progress and freedom against landowners and townspeople; Marxists discover that revolution can be based on the peasants without waiting for the development of an industrial proletariat.<sup>39</sup>

The conception of progress is itself an idea shaped by and expressive of one culture or civilization, that of the recent West.<sup>40</sup> What Toynbee and others have called the "Westernization" of the world may be the spread of only parts of the ideas associated in the West with the word "progress." Not without investigation can it be safely assumed that the spread of Western ideas from cities carries into the countryside a new and Western value system emphasizing hard work, enterprise, a favorable view of social change and a central faith in material prosperity. In the cases of some of the peoples affected by modern urbanization these values may be already present. In other cases the apparent spread of progress may turn out, on closer examination, to be a return to ancient values different from those of the West. Nationalistic movements are in

(38) For further discussion of these concepts of "mood," "aspiration" and "policy" as they might figure in community studies, see Redfield, The Little Community, chapter on "Little Community as a History."

(39) David Mitran, Marx and the Peasants.

(40) See A. L. Kroeber, Anthropology, Secs. 127, 128; Milton Singer, Shame Cultures and Guilt Cultures, for an examination of some of the evidence on this point for American Indian cultures. Also see Redfield, A Village that Chose Progress, esp. Chapter 8, "Chan Kom, Its Ethos and Success." Recent material on cross-cultural comparisons of value systems will be found in Daryll Forde (ed.), African Worlds, and in the forthcoming publications of the Harvard Values Study Project directed by Clyde Kluckhohn.

part a nostalgic turning back to local traditional life. We shall understand better the varieties and complexities of the relations today between city and country as we compare the values and world views of the modernizing ideologies, and those of the Little and Great Traditions of the cultures and civilizations that are affected by the modern West. It may be that such studies<sup>41</sup> will discover greater "ambivalence" in the mood to modernize than we, here in the West, acknowledge; that the progressive spirit of Asia and Africa is not simply a decision to walk the road of progressive convictions that we have traversed, but rather in significant part an effort of the "backward" peoples to recover from their disruptive encounters with the West by returning to the "sacred centers" of their ancient indigenous civilizations.

---

(41) Several such studies have been made. See, e.g., Paul Mus, Viet-Nam, l'histoire d'une guerre, Paris, 1952 (?); Shen-Yu Dai, Mao Tse-Tung and Confucianism, Doctoral Dissertation, University of Pennsylvania, 1952; E. Sarkisyanz, Russian Weltanschauung and Islamic and Buddhist Messianism, Doctoral Dissertation, University of Chicago, 1953 (?). V. Barnouw, "The Changing Character of a Hindu Festival," American Anthropologist, February, 1954.

# 17 The City in Informational Society

Toshio Sanuki

## INFORMATIONAL DIFFERENCES

“**T**he 21st Century is coming within the firing range.” Thanks to the continuing increase of technical innovation, level of accumulated knowledge, and level of income, we are in a better position than ever before to realize our desires. In conventional society, only people in specialized fields or members of a privileged class have enjoyed the fruits of new technical developments. The history of mankind eloquently attests to this fact.

However, in the second half of the 20th Century, technical progress has suddenly made a society in which these fruits are available to all. There are three principal causes for this development. The first is the great jump in the level of personal income. The increase in per capita income from \$113 in 1950 to \$1,200 in 1968 in Japan is a growth of 1,060%.

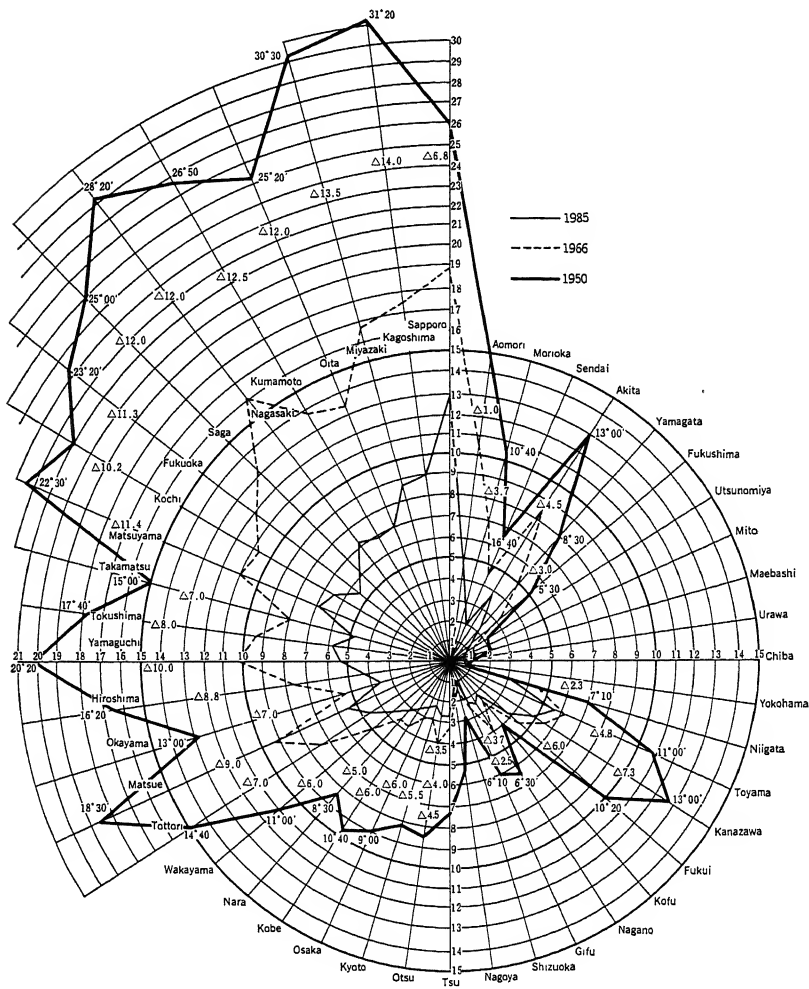
The second cause is the transportation revolution. In the beginning of the Meiji era, travel time between Tokyo and Osaka was 360 hours, but the Tokaido Line now takes 20 hours to cover the same distance. In addition, the same distance can now be covered in three hours 10 minutes with the New Tokaido Line. In other words, it now takes 1/120th of the time it formerly took to travel the same distance from Tokyo to Osaka. This is true of other places as well. Travel time from Tokyo to all the principal cities of Japan has been tremendously shortened.

The third cause is the information revolution. The widespread use of the telephone, radio, and television has made the diffusion of information throughout Japan virtually instantaneous. One hundred million people can keep up with the constantly changing news of Japan and the world while sitting at home.

These are the three principal reasons why Japan and the world have become so incredibly small. The transportation revolution has effectively subjugated physical space to the point where, for the wealthy at least, freedom of movement is an everyday reality. For the poor, however, the economic realities of transportation are still quite fettering, but the constant rise in the level of income is overcoming even this once powerful barrier. In societies with

Reprinted from *Area Development in Japan*, no. 3 (1970), pp. 9-23, with permission of the Japan Center for Area Development Research.

Graph 1 Changes in Travel Time from Tokyo to Other Cities in Japan  
— Via National Railway —

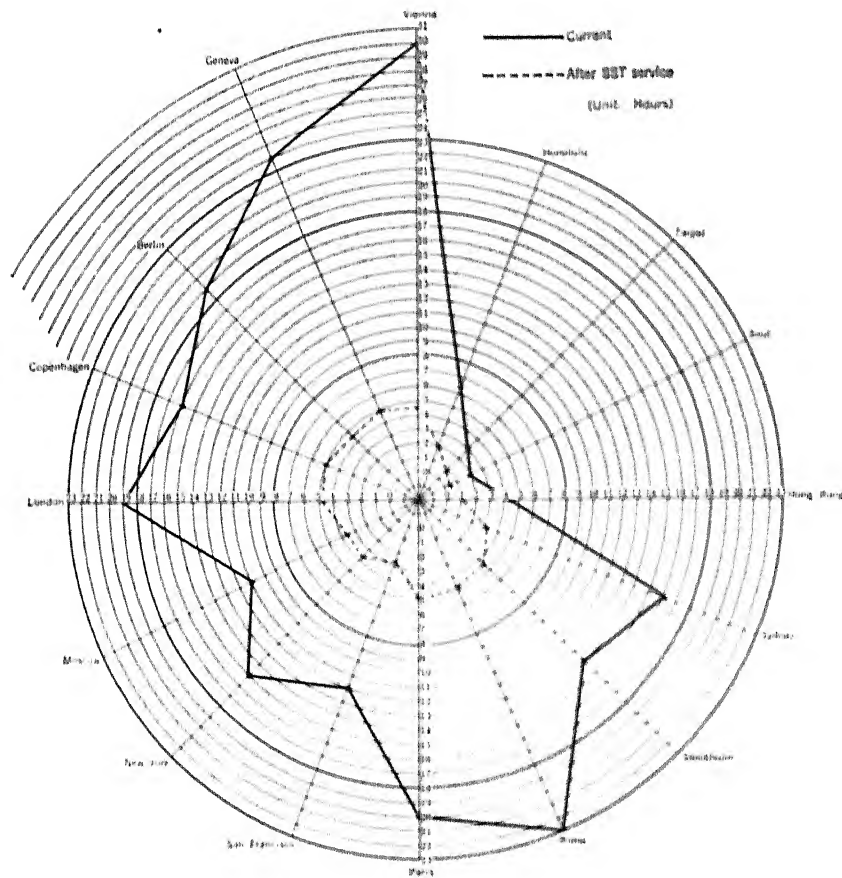


Source: "Clunging Japan," by Toshio Sanuki

a high income level, economic distance is no longer a decisive factor. We now measure distance in terms of time, not money. As shown in Graphs 1 and 2, the time-distance ratio between Tokyo and the major cities of Japan has undergone a dramatic change, and will continue to decrease in the future.

However, even more than the transportation revolution, we should note the

Graph 2 Travel Time from Tokyo to Major World Cities



dramatic changes which the information revolution is bringing about in the time-distance between specific major capitals and specific cities of the world. In this respect, has the time-distance between all these major cities been reduced to zero? To ask the question is to beg the answer. The information revolution is reducing time-distance between specific major cities at a very rapid rate. Therefore, one of the problems of the new informational society is that, just as there are presently very large problems of personal income differences, only a limited number of cities are becoming centers of information, with many other cities lagging behind. This creates the possibility of differences in the level of information sophistication.

Table 1 Scale of Agglomerated Central Management Functions

(62 Cities=100)

|        | <i>Economic<br/>Function</i> | <i>Administrative<br/>Function</i> | <i>Cultural &amp; Social<br/>Function</i> | <i>Composite<br/>Index</i> |
|--------|------------------------------|------------------------------------|---|----------------------------|
| Tokyo  | 44.2%<br>(100.0)             | 43.6%<br>(100.0)                   | 39.9%<br>(100.0)                          | 42.6%<br>(100.0)           |
| Osaka  | 17.6%<br>( 39.8)             | 5.8%<br>( 13.3)                    | 11.3%<br>( 28.3)                          | 11.6%<br>( 27.2)           |
| Nagoya | 6.0%<br>( 13.6)              | 4.6%<br>( 10.5)                    | 5.9%<br>( 14.8)                           | 5.5%<br>( 12.9)            |

#### URBAN POLICY IN RESPONSE TO THE INFORMATIONAL SOCIETY

As I have elsewhere outlined my views on the relationship between the rise in the level of income and the progress of urbanization,<sup>(1)</sup> I shall not go into this question here. However, the rising level of income and the progress of informational society are very closely related. As is clearly shown in Graph 3, there is a proportional relationship between the level of income and the level of informational society. In other words, where the level of income is the lowest, as in Kagoshima, the index of informationalization is also the lowest.

On the other end of the scale is Tokyo with the highest level of income, whose index of informationalization is 213.6, or 4.1 times greater than that of Kagoshima. As the income level of Tokyo is only 2.9 times greater than that of Kagoshima, we can clearly see that the difference in the level of informational society far exceeds this proportion. Not only this, but the difference in informational level between Tokyo, the highest, and Osaka, the second highest, is very large. Even more striking than this is the unexpectedly large gap between Tokyo and Nagoya (Aichi on the Graph). This indicates the close relationship of information level with the scale of central management functions which the city performs. Table 1 clearly shows that informational functions and central management functions are basically two sides of the same coin.

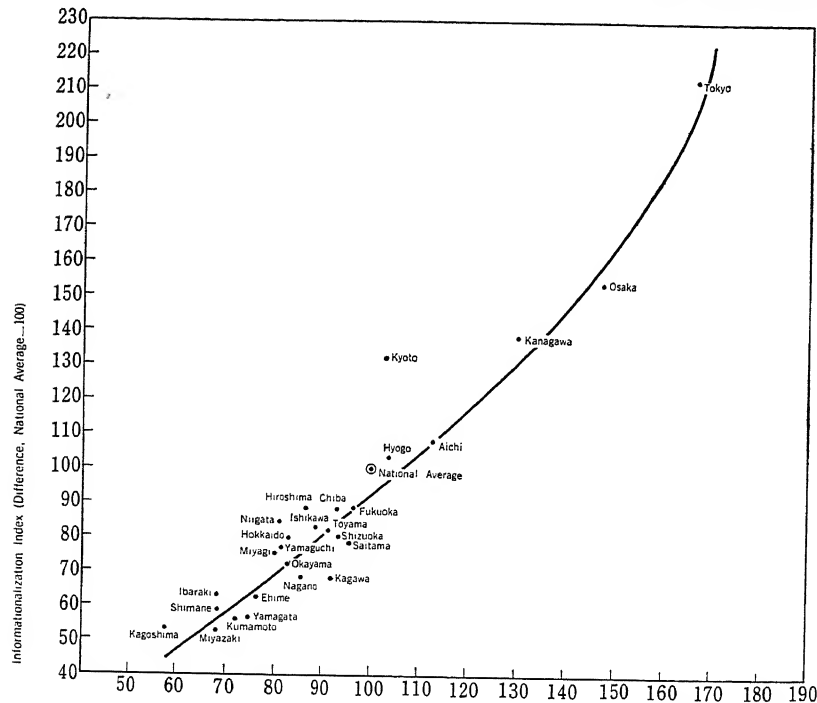
As income rises, people move to the cities. There is a proportional relationship between the level of income and the density of population in DID.<sup>(2)</sup> Does society become more informational as urbanization advances? Graph indicates that the higher the level of urbanization, the higher the level of informationalization. Above an urbanization level of 70%, informationalization advances very rapidly, and above 90% the rate becomes even greater.

The more urbanization advances, the higher the population density and the density of all types of economic activity become. Therefore, it cannot be

(1) Toshio Sanuki, *Chitki-Kaitatsu to Kinyu* (Regional Development and Finance), Zenkoku Chiho Ginko Kyokai, Tokyo, 1968.

(2) Densely inhabited districts in the Japanese census are those districts with a density of more than 4,000 per square kilometer and with a population of at least 5,000

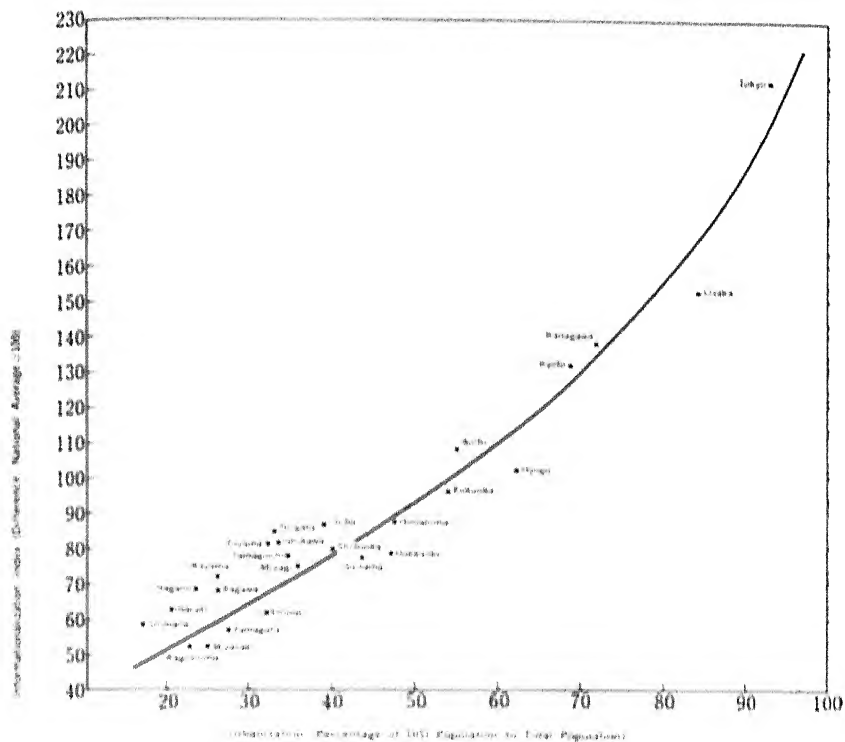
Graph 3 Correlation Between Income and Level of Informationalization



wrong to assume that the density of information will also rise accordingly. When we turn to present day figures in trying to prove this supposition (see Graph 5), we can see that as urbanization increases, informationalization increases at an extraordinarily higher rate. Looking at Graph 5 (whose vertical axis is logarithmically graduated) we can see that Tokyo, whose income level is only 2.9 times that of Kagoshima, has an information density which is 80 times that of Kagoshima. We cannot help but marvel at this extraordinary difference. It is perfectly clear that there is a large disproportionate relationship between urbanization and informationalization. In an age of urbanization the amount of information flow in society, government, law, administration, and indeed all fields increases very rapidly. Moreover, because these fields are closely interrelated and interact with one another, the cumulative effect is to expand information flow at a very rapid pace.

What kind of relationship is there between the scale of the amount of information and the density of information? They are correlative, as illustrated on Graph 6. The cities can be categorized into four groups: the leading group, the advanced group, the middle and the last group. Hokkaido is the only exception in which the amount of information and the density of information are not proportional, because with 21.2% of the land area of Japan and 12.6%

Graph 4 Correlation Between Urbanization and Informationalization

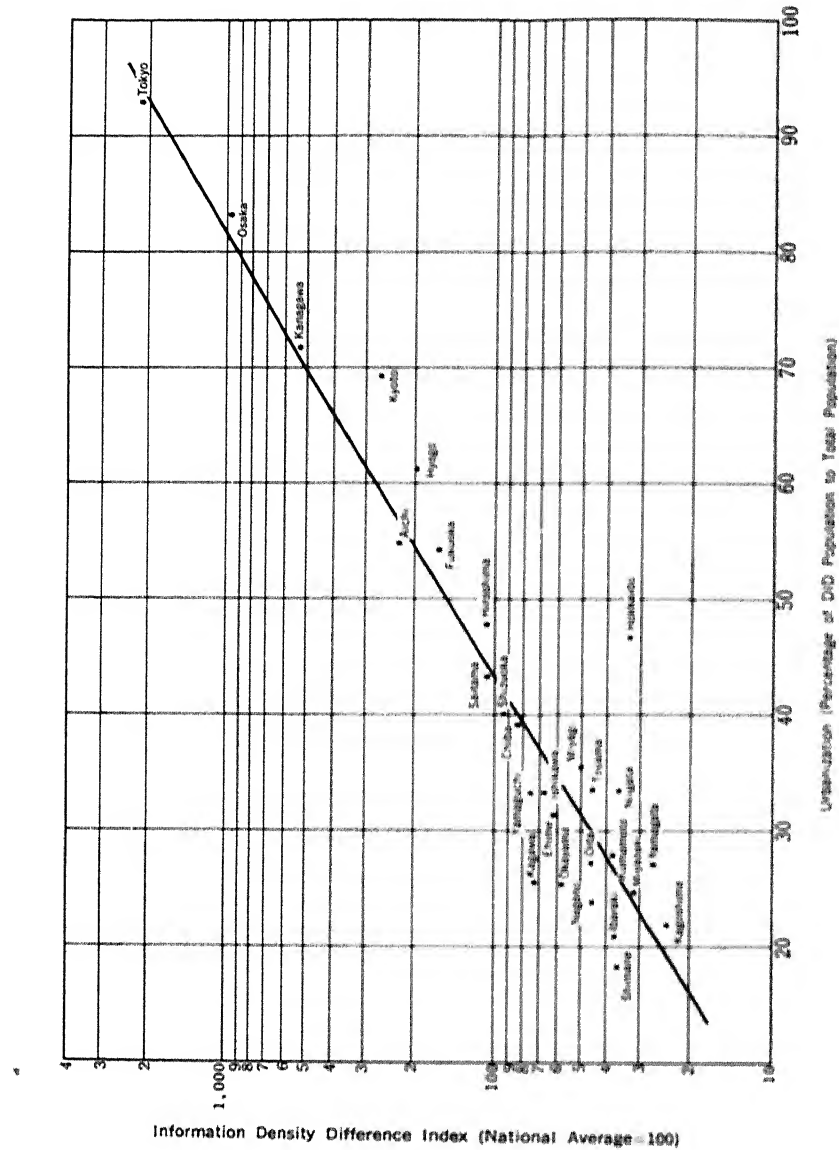


of the habitable land area, it is Japan's only region with the possibility for completely new development. Therefore, as urbanization advances, population concentrates in the cities, economic and government functions concentrate in the cities, and the scale of this agglomeration of central management functions produces vast amounts of information. If we trace the correlation between the index of informationalization and the density of information, it will appear as on Graph 7.

From the beginning of the Meiji era, when 86% of Japan's labor force was devoted to agriculture, through Japan's ages of light and heavy industrialization to its present age of urbanization, the tempo of information exchange and accumulation has kept increasing its pace accordingly. Now we are entering the post-industrial age. Assuming that this is a society in which the knowledge industries control all other industries, we can also assume that the coming of informational society to all regions and cities will be the principal factor in producing revolutionary changes. It will, therefore, be necessary for us to create new urban policies to develop a regional society in which the production,



Graph 5 Correlation Between Urbanization and Information Density



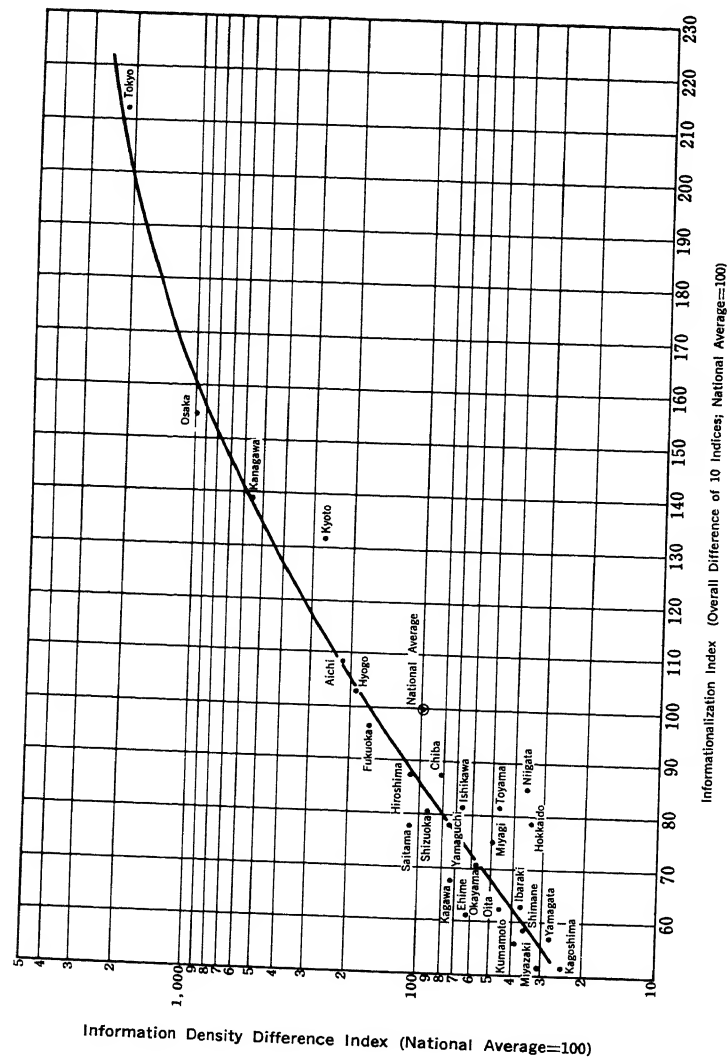
exchange, buying and selling, and consumption of information have become of paramount importance. In order to make such new policies, we must first get an idea of what the information flow pattern between the major cities of Japan is.

This image shows a dark, textured vertical band, likely a book cover or binding. A lighter, textured vertical strip runs down the center, possibly indicating a hinge or a different material. The overall appearance is grainy and aged.



HUNT LIBRARY  
CARNEGIE-MELLON UNIVERSITY  
PITTSBURGH, PENNSYLVANIA 15213

Graph 7 Correlation Between Informationalization and Information Density



coefficient of dispatched messages, we can see that Tokyo occupies 33.7%, Osaka 28.5%, Nagoya 12.9%. The three largest metropolitan areas thus occupy 75.1% of the network, with all the rest together occupying less than 25%. The fact of the concentration of information in these large cities is obvious.

This being the case, two problems in regards to the formation of an information network remain. The first is the formation of main and branch information networks between the principal cities. The second, I think, is what form should

the information network take within the large metropolitan area to facilitate information flow. As regards the former, I do not think there are terribly difficult problems. The situation here is very similar to that of the highway network linking the areas. However, the formation of an information network within the large metropolitan areas presents precisely the same problems as those encountered with the transportation system. Very difficult problems must be solved in this area.

Let us leave the analysis of the information flow pattern within the metropolitan area to another occasion, and instead concentrate our attention on the information flow pattern between the major cities of Japan.

I have already pointed out that the Tokyo, Osaka, and Nagoya metropolitan areas account for more than 70% of the information flow in Japan. If we then look at the information flow pattern out of Sapporo, we see that 79.5% is directed to Tokyo. Adding to this the 9.5% which is directed to Osaka, it becomes roughly 90%. We can, therefore, say that, as far as the information flow pattern is concerned, Sapporo is really inside the sphere of influence of the Tokyo metropolitan area. Graph 8 illustrates this fact.

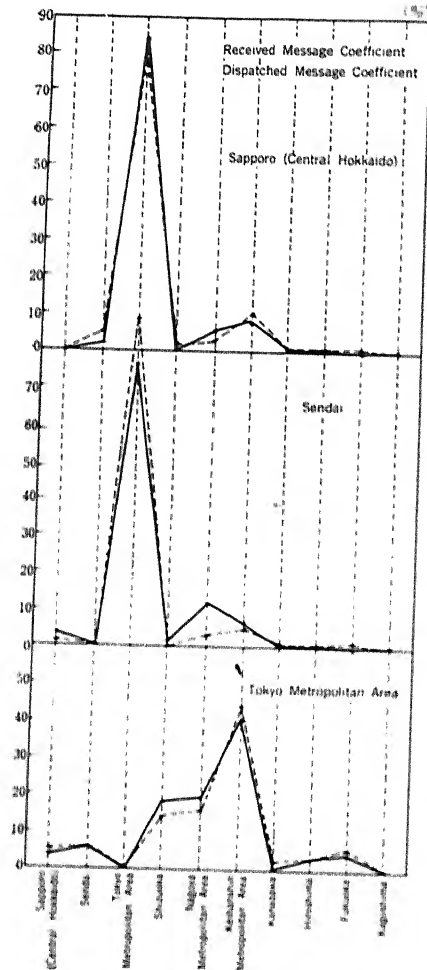
What about Sendai? Its dependence on Tokyo is even greater, as 88% of its information output is directed to Tokyo. Adding to this the information sent to Osaka, it becomes 93.5%. Opposed to this pattern is Tokyo, which sends 44.2% of its information to Osaka, 16.6% to Nagoya, 15.2% to Shizuoka, 6.1% to Sendai, 6% to Fukuoka, 5.4% to Sapporo, 3.4% to Hiroshima, 1.9% to Kanazawa, and 6.5% to Kagoshima. From this we can easily see that Tokyo is the center from which directives are issued to all the principal cities of Japan, thus amply fulfilling its role as the management center.

Let us take a look at Shizuoka. It sends 76% of its information to Tokyo. Although Nagoya is its closest neighbor, it still sends more information to the Osaka area. Shizuoka is far more dependent on Tokyo for its information flow than is estimated by the migration of population, movement of passengers on the New Tokaido Line, and movement of goods between Shizuoka and Tokyo.

Nagoya receives the same amount of information from Tokyo and Osaka and replies in like amounts. In other words, Nagoya is equally dependent on Tokyo and Osaka. Excluding Shizuoka, the amount of information exchange with other cities is almost nil. Looking at Nagoya from the information point of view, it cannot be called one of the informational centers of Japan. We can say that only the two largest cities of Tokyo and Osaka are true information centers.

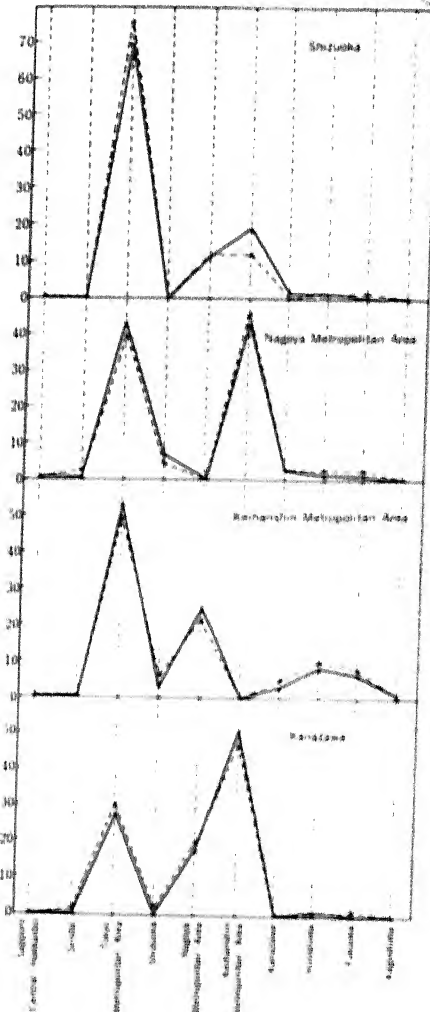
The amount of information sent from Osaka to Tokyo is 51.5% of the total information outflow from Osaka, and that to Nagoya is 21%. Osaka also sends a fair amount of information to the cities of western Japan, Hiroshima, Fukuoka, etc., as well as to Kanazawa. We can, therefore, presume that Osaka is the central management function center for western Japan as well as for part of the Japan Sea Coast region.

Graph 8-1 Information Flow Pattern Between Principal Cities



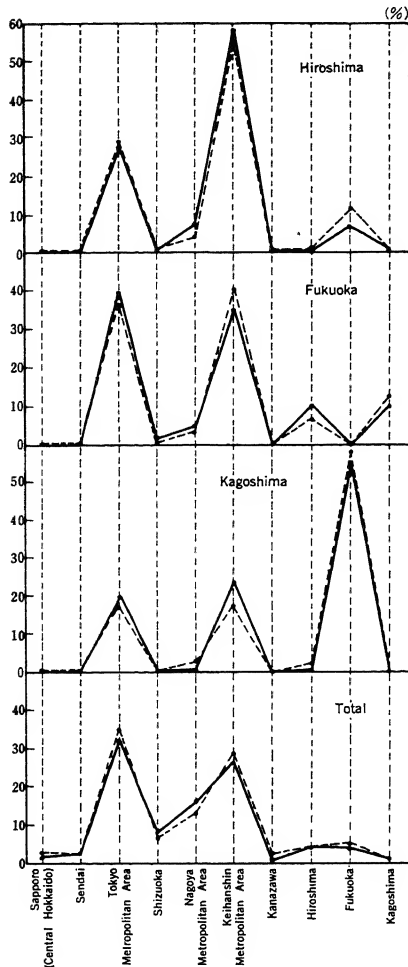
Note: Received and dispatched message were calculated on the basis of "Ito" model developed by Professor Zenichi Ito, Tokyo Women's University

Graph 8-2 Information Flow Pattern Between Principal Cities

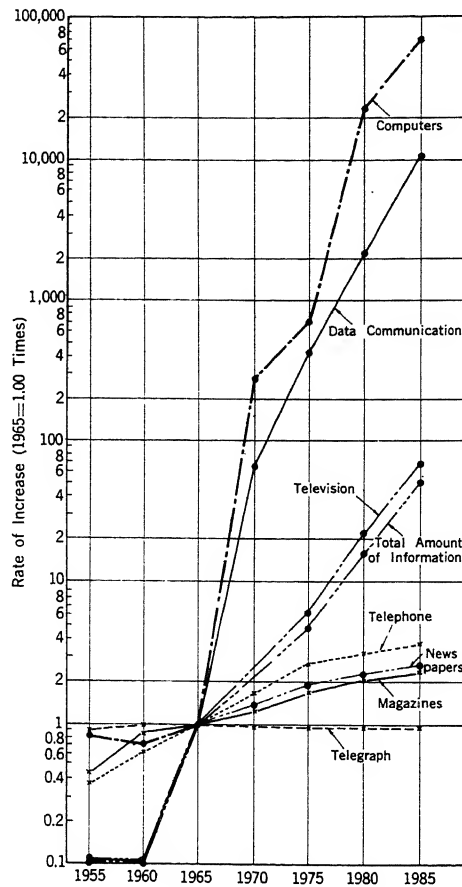


What about Kanazawa? Kanazawa is fully dependent on Osaka. Known as "Little Kyoto," it has had a close relationship with the Osaka-Kyoto-Kobe area since Tokugawa times, and this tendency is even stronger today. Kanazawa sends 48.7% of its information to Osaka, 29.1% to Tokyo and 18.4% to Nagoya. It stands roughly middle way between the extremes of Tokyo and Nagoshima.

Graph 8-3 Information Flow Pattern Between Principal Cities



Graph 9 Comparison of the Amount of Information Carried by Various Information Transmitters



When we look at Hiroshima, we see that it is overwhelmingly dependent on Osaka. Fifty-five and four-tenths percent of its information goes to Osaka, 27% to Tokyo, 11.5% to Fukuoka, and an insignificant amount to Nagoya.

Will this information flow pattern continue indefinitely into the future? As the information flow pattern is intimately related to the transportation system, their mutual interdependence is assured. We can, therefore, assume that when the extended New Tokaido Line is completed and opened to traffic there will

be large changes in these patterns in future. Also, the pattern in which Tokyo is the center of information agglomeration can be expected to rapidly strengthen as the transportation revolution advances.

However, opposed to this is the quite different information flow pattern Fukuoka. The amount of information it sends to Osaka, 40.2%, is about the same level as that it sends to Tokyo, 37.5% (Graph 8). However, it sends the relatively large amount of 11.6% to Kagoshima, and we can, therefore, say that it has the characteristics of an information center for Kyushu. If we look at this relationship in reverse, we see that Kagoshima sends 58.8% of its information to Fukuoka. In other words, it is heavily dependent on Fukuoka. If we now compare the amount of information sent from Kagoshima to Tokyo, 18.8% and to Osaka, 18.0%, we see that it is more dependent on Tokyo. We must especially note that although in terms of physical distance Osaka is far closer yet the pattern of greater dependence on Tokyo remains clear.

The conclusion which we draw from this analysis is that Tokyo and Osaka are true information centers, but Nagoya is not. Also, other regional cities like Nagoya, are not able to be information centers. They are only able to function as relay stations for the information centers of Tokyo and Osaka. The information flow pattern is intimately related to the formation of the transportation network. From the point of view of informational functions, the form of the transportation system will greatly transform the characteristics of the important regional cities.

### *THE CITY IN INFORMATIONAL SOCIETY*

The distribution pattern of cities will be greatly transformed by the form of the transportation network and the future form of the information network.

In the most advanced regions of Japan industrial society has already been superseded by post-industrial society. Since 1960, in Tokyo, for every additional 100 blue collar workers there have been 466 additional white collar workers. At present, industries are moving out of the Tokyo area to less developed areas. In other words, Tokyo has already become the post-industrial society.

In the post-industrial society the agglomeration of knowledge and information will absolutely determine the nature of future cities. The largest metropolitan areas are already rushing forward into the knowledge and informational society, and the other cities I mentioned previously are following the same path at greater or lesser speeds. In contrast to this is the fact that small and medium size cities in the underdeveloped regions have been entering industrial society since the mid 60's. The cities of Japan show a pattern of being at various levels of development. If we now try to formulate one standard urban policy without taking into account these differences in development stage, we can only increase the possibility of making serious mistakes.

However, the time lag between these various levels of development is rapidly diminishing. Can the small and medium size cities become full members of the informational, post-industrial society? A weeding out process of conventional industries in small and medium size cities will take place as these industries lose their ability to compete successfully with similar industries in the developing countries. It is no longer possible to depend on the old industries. A period of movement and realignment toward cities in which new knowledge and information are necessary is coming soon.

With this movement toward an informational society, it has become an absolute condition that future cities must reach a certain level of knowledge-information agglomeration. In places where this economy of scale is not present, the need for knowledge and information is disappearing. Farming villages and small- and medium-size cities are not the places where information and knowledge are produced, distributed or consumed. It is only in large cities where the economy of scale can be used that information is produced, distributed and consumed.

For the future city, the most important condition for coping successfully with informational society will be extreme largeness of scale. Before small- and medium-size cities have emerged from an agricultural society into an industrial one, Japan in general has entered the informational society. In present-day terms, small- and medium-size cities will lose the chance for development as they no longer perform vital functions. There is the strong possibility that the old functions which they are presently performing will rapidly be scrapped. Accordingly, small- and medium-size cities will be naturally weeded out even faster by the advance to an informational society level than by the earlier process of industrialization.

In order to verify these contentions, let us look at Graph 9 to see in which communication areas informational society is being propelled forward. It is calculated that the amount of information in 1985 will be 52.5 times greater than it was in 1965. If we then look at the means of information transmission, we see that the telegraph is diminishing, and that the growth rate of magazines, newspapers, and books, *i.e.*, written information, is relatively poor. As opposed to this is television, which produces and circulates pattern and color information, whose growth rate is relatively high. Beyond this is the explosively growing data and computer information field which will play a leading role in the future society. With this in mind, we can see that the number of computers in use will grow rapidly throughout the major countries of the world. Among them, we can expect that Japan will have the highest growth rate. We can expect that Japan will be the nation most rapidly moving into a high density informational society.

Where will this development take place? It will take place in the very largest metropolitan areas. These are the places where urban policy will be of the greatest importance. This development will occur most strikingly in the



two largest agglomerations of Tokyo and Osaka, but not in Nagoya or the regional cities. It is, therefore, easy to solve the problems of information between the principal cities of Japan, but the problems of creating an information transmission system within the largest metropolitan areas and an effective government policy to handle the new informational age are very crucial and difficult.

We are now at the stage where we must turn our attention to the problem of making a new theory of information functions in order to formulate future urban policy. Only then, having been able to solve these problems, can we begin to create the new informational society city of the future.

# 18 The Analysis of "Over-Urbanization"

N. V. Sovani

THE rapid rate of urbanization in the underdeveloped areas of the world during the last two decades has attracted widespread attention and has evoked quite a considerable body of analytical writing from social scientists. Out of this discussion a broad structure of analysis seems to be taking shape. I would conveniently describe it as analysis bottomed on the concept of "over-urbanization." The ground work is laid by defining "over-urbanization," and the upper structure is provided by the discussion of its causes and consequences. Asia, for example, is said to be over-urbanized at present, in the sense that "at comparable levels of urbanization, the developed countries of today had a correspondingly greater proportion of their labor force engaged in non-agricultural occupations."<sup>1</sup> This over-urbanization is supposed to have come about because rural migrants have been "pushed" rather than "pulled" into the urban areas in these countries, as a result of great and mounting population pressure in the rural areas. In over-urbanized countries "urban misery and rural poverty exist side by side with the result that the city can hardly be called 'dynamic,' as social historians of developed countries generally described the process of urbanization."<sup>2</sup> The purpose of this paper is to examine critically these three propositions. The first, dealing with the definition of "over-urbanization," is mainly methodological, while the other two are empirical in character. I will consider them in the same order as above.

## I. Definition of "Over-Urbanization"

In defining over-urbanization, two indices are being related to one another, the percentage of population living in urban areas, and the distribution of the total labor force in the country as between agricultural and non-agricultural occupations. The first is a

<sup>1</sup> *Urbanization in Asia and the Far East*, Proceedings of the Joint UN/UNESCO Seminar, Bangkok, 8-18 August (Calcutta: UNESCO Research Center on the Social Implications of Industrialization in Southern Asia, 1957), p. 8.

<sup>2</sup> *Ibid.*, p. 10.

Reprinted from *Economic Development and Cultural Change*, vol. 12, no. 2 (January 1964), pp. 113-122, with permission of the author and The University of Chicago Press. © 1964 by the University of Chicago.

spatial index without being an occupational one, and the second is the opposite. As modern urbanization is associated with industrialization, we may generally agree that there is justification for such a comparison.

The next question is what should be regarded as the normal relationship between the two indices so compared. In this context, two kinds of norms have been suggested. One is based on cross-section analysis of data for a large number of countries in the world, in and near 1950, and is proposed and used by Davis and Golden;<sup>3</sup> another based on historical analogy is suggested and used by the UNESCO Seminar Report quoted earlier.

Davis and Golden take the percentage of economically active males not engaged in agriculture and the percentage of population in cities of 100,000<sup>4</sup> and above in a large number of the countries

<sup>3</sup> Kingsley Davis and Hilda H. Golden, "Urbanization and the Development of Pre-Industrial Areas," *Economic Development and Cultural Change*, Vol. III, No. 1, October 1954.

<sup>4</sup> The index of urbanization used by Davis and Golden is the percentage of the total population of a country living in cities of 100,000 and more. In maintaining that this index is valid for comparative purposes, they observe:

"Actually, since there is a certain regularity about the pyramid of cities by size, the proportion in any major size-class tends to bear a systematic relation to the proportion in other size-classes. Thus the percentage of a population living in places above 100,000 has a ratio to the percentage in places above 5,000 which is roughly similar from one country to another. An index of urbanization is therefore quite feasible for comparative purposes (*ibid.*, p. 7, fn. 1)."

Unfortunately, this is not so. A tabulation of the ratio between the percentages of population in places of 5,000 and more (P1) and those of population in cities of 100,000 and more (P2) for countries from the 1952 Demographic Year Book of the UN reveals the following:

| <i>Range of the<br/>ratio P2/P1</i> | <i>Number of<br/>countries</i> |
|-------------------------------------|--------------------------------|
| .10 to .30                          | 8                              |
| .31 to .40                          | 11                             |
| .41 to .50                          | 14                             |
| .51 to .60                          | 12                             |
| .61 to .70                          | 7                              |
| .10 to .70                          | 52                             |

Thus the validity of the index of urbanization used by Davis and Golden is rather questionable. Their observation is probably based on the rank-size rule or the Pareto distribution that characterizes the sizes of cities in many countries. In the first place, this rule is not universal, and in several countries, particularly those dominated by single large or primate cities, the

in the world in and near 1950 and find a correlation coefficient of 0.86 between the degree of industrialization and the degree of urbanization. When the "relationship between the two variables is represented in the form of a regression curve, certain countries are found to be off the line to a significant extent."<sup>5</sup> Egypt, Greece, Korea, and possibly Lebanon are found to be off the line, i.e., in them the degree of urbanization is much more than would be expected from the level of industrialization they have achieved.<sup>6</sup>

This criterion has been derived from a correlation found to exist between two variables at a point or over a short stretch of time. It may be considered, as a first approximation, a fairly useful criterion for identifying cases of urbanization that do not conform to the broad pattern found to be prevailing at or around a point of time and which therefore need special attention or study. But before proceeding further, it would be necessary to examine the correlation itself, to test its stability through different stages of industrialization and through time in the different countries. For the first, we can examine how it works for subgroups of countries that are in similar stages of industrialization. For the second, we can examine whether the cross-section analysis is also borne out by the time-series analysis.

As a first step towards this I tried to work out the correlation coefficient for countries for which data on urbanization and occupational distribution were available for the same year and in the period from 1946 to 1951 in the UN's Demographic Year Books of 1952 and 1956. These data were available for a total of 41 countries, excluding countries or geographical units that had no

size distribution of cities does not conform to it. Secondly, even if the rule were universal, the slope of the Pareto curve will differ from country to country, and the ratio between the percentages of population in different size-classes of towns will therefore differ. The value of  $\alpha$  ranges from 0.93 to 1.59 in the case of six countries for which Singer gives the values. Cf. H. W. Singer, "Courbes des Populations: A Parallel to Pareto's Law," *Economic Journal*, 1936, pp. 254-63. The necessary and sufficient condition for the Davis-Golden observation to be valid would be (a) that the rank-size rule is universal, and (b) that the slope of the distribution curve is either the same everywhere or varies only slightly as between different countries.

I will go no further here than to indicate the questionable assumptions underlying one of the variables used by Davis and Golden, and in the rest of the paper, I will treat their index as valid.

<sup>5</sup> Davis and Golden, *op. cit.*, p. 8.

<sup>6</sup> Correspondingly, on the other side of the regression line there would be cases of "under-urbanization." If one is considered abnormal in some sense those on the other side of the regression line also are abnormal.

cities of 100,000 and more. The correlation coefficient was found to be 0.70.<sup>7</sup> In order to see how the correlation holds in highly industrialized countries as against the rest, I worked out the correlation coefficient for a group consisting of the U.S., Canada, and 15 European countries for which data were available, and another for the remaining group of 24 countries in my original list. Surprisingly, the correlation coefficient between urbanization and industrialization for the group of highly industrialized countries was 0.395, and that for the remainder was 0.85. These results indicate that the association between the two variables is much more close in the underdeveloped countries than in the highly industrialized countries or, by implication, that the pace of urbanization in the underdeveloped countries is much more closely dependent on the pace of industrialization than in the highly industrialized areas. This flies in the face of the entire over-urbanization thesis, at least in the way it has been formulated up to now.

Yet the results obtained above make sense, because what they bring to light is the non-homogeneous character of the two groups of countries and the invalidity of deriving correlation coefficients for all countries together in one lump. An analogy should make this clear. That there is a strong correlation between the height and weight of human beings is well known. But if this correlation is considered for persons of different age groups it will be certainly far stronger in the age groups below 20 than in those above 20 years, in which persons have approached or are approaching the asymptotic limit of their height. The same is true of countries at advanced and early stages of industrialization and urbanization.

This is supported by a similar correlation analysis done for U.S., Canada, and eleven Western European countries for the year 1891,<sup>8</sup> which gives a correlation coefficient of 0.84. The occupational data are for the entire labor force and not for males alone as in the analysis of recent data. But that would mean that the correlation coefficient here is slightly smaller than what it would be if data for males alone were used. The indication, however, is clearly that in

<sup>7</sup> The higher value of  $r$  and  $r^2$  that Davis and Golden obtained might have been due to the use of data for more countries than I have been able to get together.

<sup>8</sup> Urban population data are from A. Weber, *The Growth of Cities in the Nineteenth Century, A Study in Statistics* (New York: Columbia University, 1899). Labor force data are from S. Kuznets, "Quantitative Aspects of the Economic Growth of Nations. II. Industrial Distribution of National Product and Labour Force," *Economic Development and Cultural Change*, Vol. V, No. 4, Part II, July 1957.

the earlier stage of industrialization and urbanization in these countries, the correlation was much stronger than now, when both the processes have gone much further.

I tried to test this further by studying time-series data for these variables in England and Wales, the U.S., Canada, France, and Sweden. Data for other countries was not easily available to me, but that for the few countries I could study gave significant results. When the two variables are plotted for each country separately and compared over periods varying from 80 to 100 years, the two curves are found to be broadly of similar shape but to differ considerably with regard to the distance between them at different times, as well as in the way they develop or grow with time.

The conclusion that emerges from this is that the correlation worked out by Davis and Golden varies at different stages of industrialization and is not stable through time. They are therefore clearly in the wrong when they apply their correlation historically to Egypt. They calculate the expected levels of urbanization in Egypt in 1907, 1917, 1927, 1937, and 1947 from the regression equation and compare the expected to the actual levels of urbanization, concluding that there was "over-urbanization" in Egypt at these points in time. If the case of Egypt in 1947 is judged from the regression equation worked out by me for the 24 countries, outside Europe and excluding the U.S. and Canada, it is found to conform very much to the general pattern.

The other criterion based on historical experience elevates the course of urbanization and industrialization in some developed countries, namely, the U.S., France, Germany, and Canada, into a norm and regards the proportions subsisting between the two indices at different times in their evolution as a measure of normality. If the proportion of population living in cities of 100,000 and above is used as an index of urbanization,

... "it can be said that roughly one in twelve in Asia is a city-dweller as against one in eight in the world as a whole, approximately one in three in North America, and one in five in Europe (including U.S.S.R.).

Such a degree of urbanization is associated in Asia with a degree of industrialization corresponding to 30 per cent of the labor force engaged in non-agricultural activities. At comparable periods of urbanization levels the United States (1850's), France (1860's), Germany (1880's), and Canada (1890's) had roughly 55 per cent of their labor force engaged

in non-agricultural occupations. Thus Asia can be said to be comparatively over-urbanized in relation to its degree of economic development."<sup>9</sup>

An implicit assumption here is that the course of industrialization and urbanization in all countries should radiate more or less in close conformity to the path taken by them in the four countries mentioned, if it is not to be classed as abnormal. When stated in this way, few will be disposed to regard it as in any way decisive or valid. The only reason for regarding the situation in a few developed countries as the norm for the rest of the world seems to be nothing better than the fact that they are today developed economies. But even if we judge other developed countries at some period of their development, we will find that they did not conform to this standard. For example, when in 1895 the degree of urbanization in Sweden was comparable to that of Asia today (8.2 per cent in cities of 100,000 or more), the proportion of the labor force in non-agricultural occupations there was less than 45 per cent. Even in 1910, though urbanization had increased slightly to 9.3 per cent, this proportion was only 51 per cent. Conversely, in Switzerland, though the proportion of the labor force in non-agricultural occupations was 60 per cent in 1888, there was no city with a population of 100,000 or more in the entire country at that time. In fact, if we logically pursue the analysis based on this norm, the whole of South and Central America would have to be classified as over-urbanized, and for that matter, the whole of Africa and so too the world ! One can turn around and say that, compared to the world outside, these four countries are really over-industrialized or under-urbanized with equal justification !

<sup>9</sup> UNESCO Seminar Report, *op. cit.*, p. 133. Not that it is materially important for the subsequent argument, but to set the record straight this statement appears to be true of Canada and Germany in the respective periods noted, but not of the U.S. and France. In the U.S. the comparable level of urbanization was reached in the 1850's, but the proportion of the labor force in non-agricultural occupations varied between 35 and 46 per cent between 1850 and 1870; see Colin Clark, *Conditions of Economic Progress*, 2nd ed. (London, 1951), p. 404. In France the comparable level of urbanization was attained in the 1860's, but the proportion of the labor force in non-agricultural occupations was around 48 per cent between 1856 and 1876; see F. Simiand, *La Salaire, L'Evolution Sociale et La Monnaie* (Paris, 1932), quoted by Kuznets, *op. cit.*

It may be noted, if it has not already become obvious, that the two criteria of over-urbanization discussed above conflict with one another. Eight to nine per cent of the population living in cities of 100,000 and more will be associated, in the Davis-Golden regression, with 30.5 per cent of the labor force engaged in non-agricultural occupations, and not 55 per cent as under the second criterion; conversely, 50-55 per cent of the labor force would be associated with about 18 to 20 per cent of the population in cities of 100,000 or more. Several cases can be cited where over-urbanization exists according to the second criterion but not according to the first.

It is surely unnecessary to further labor the obvious—that the definitions of "over-urbanization" developed so far are chimerical and so unusable.

## II. *Causes of "Over-Urbanization"*

From the criteria of over-urbanization, let us now pass to the alleged causes of this phenomenon. The main one, according to the current analysis, is the pressure of population on land in the rural areas in these countries. Economic pressure or "push" in the countryside mounts continuously and pushes out people to the cities in search of employment and livelihood. The rural-urban migration that leads to over-urbanization is mainly a consequence of this "push" from the countryside, rather than the demand for labor by developing economic activity in the towns and cities, or what is called their "pull." Consequently, these migrants can only get employment in activities with very low productivity or swell the ranks of the unemployed.

"Thus the recent rapid rate of urbanization visible in Asian countries does not bespeak of a corresponding growth of industry but a shift of people from low productive agricultural employment to yet another section marked by low productivity employment, namely, handicraft production, retail trading, domestic services in urban areas."<sup>10</sup>

This statement describes only a part of the reality, and its suggestive implications are darker far than they are in actuality. It is true that migrants to towns are absorbed in low productivity employment, but even so, this urban employment is found to be by

<sup>10</sup> UNESCO Seminar Report, *op. cit.*, p. 133.



and large more productive than the pre-migration rural employment of the in-migrants. The urban per capita incomes are almost universally found to be higher than per capita rural incomes in most of the countries. That there is unemployment in the urban areas is true, but can the towns and cities remain dry islands of full employment and very high labor productivity in a sea of rural unemployment and underemployment? A rate of industrial development much greater than that witnessed in these countries in recent years would not be able to change this over-all picture materially because of the enormous backlog of unemployment and underemployment in these countries. A little examination of the relevant data can easily bring this out. That urbanization in these countries does not bespeak of industrialization has been shown to be a questionable inference in Section I above.

It should next be noted that over-urbanization according to the first or the second criterion is also found in countries and areas where there is little or no pressure on land in the rural countryside. Most of the countries of Central and South America and many in Africa are in this category. There seems to be no invariant correlation between rural pressure and over-urbanization.<sup>11</sup>

Moreover, the Asian countries referred to in the quotation above have historically (under colonial rule) experienced increasing pressure on land for at least the last century. Yet it is only in the last two decades that the rate of urbanization in them has accelerated. Before 1939, most of them were noted for their very small degree of urbanization. Rural pressure, therefore, is no new factor in their situation, and if it is such a potent force in furthering urbanization, as is alleged, then why it did not formerly result in "over-" instead of "under-urbanization" has not been satisfactorily explained. On the other hand, if it is to be maintained that this pressure only reached the critical level necessary for resulting in rapid urban growth only during the last two decades, then again it will be necessary to define this critical level of rural pressure and the factors that determine it. This begs the whole question.

<sup>11</sup> Davis and Golden also did not find any correlation between the degree of urbanization in a country and the average density of population there. See *op. cit.*, p. 10. There is, however, they claim, a negative relationship between urbanization and agricultural density defined as the number of males occupied with agriculture, hunting, and forestry per square mile of cultivated land. It can be easily seen that this goes against the whole thesis of rural pressure being the main factor bringing about rapid urbanization.

Information regarding the causes of rural-urban migration in underdeveloped countries of Asia and elsewhere is very meager. An ILO report furnishes an authoritative recent analysis in this field. It concludes :

"The main push factor causing workers to leave agriculture is the lower level of incomes. In almost all countries incomes in agriculture are lower than in other sectors of the economy. The main factor determining the rate of outward movement is the expansion of employment in other occupations. It is this factor which explains the high rate of movement in recent years in the advanced countries (among which Sweden, the United States, and Canada are outstanding) and in rapidly developing countries in Latin America, the Middle East, and Africa. Although the push factors of falling incomes and underemployment in agriculture in most of the less developed countries are now very strong, they do not, in the absence of strong pull factors, suffice to cause large shifts in manpower between occupations. High rates of movement indicate rapid growth and high rates of investment, either in the economy as a whole, including agriculture, or in the industrial or urban sector, as is the case in almost all of the less developed countries which are now in the process of very rapid development."<sup>12</sup>

This indicates that the causal relationship underlying rural-urban migration is quite complicated and cannot be completely explained by the rural push factor.<sup>13</sup> The phenomenon of a rural "push" resulting in urban growth is highly questionable.

<sup>12</sup> ILO, *Why Labor Leaves the Land, A Comparative Study of the Movement of Labor out of Agriculture*, Studies and Reports, New Series No. 59 (Geneva, 1960), p. 209. The report goes on to point out that lower levels of income are a universal reason for movement. But different causes operate to reduce the level of incomes in agriculture in relation to other incomes as between the advanced and the less developed countries. In the former, labor leaves the land because it is an underprivileged sector of the economy (p. 210).

<sup>13</sup> With the available Indian data I have argued elsewhere that rural pressure exploding into urban growth is a phenomenon rarely met with in India. The migrants to urban areas form only a small percentage of the total unemployed and underemployed in the rural areas. There is also no reason to believe that the economically worse off in the rural areas only migrate. Rural economic conditions are bad, but they are not a necessary nor a sufficient condition for rural-urban migration. N. V. Sovani, "Urban Social Situation

### III. *Consequences of "Over-Urbanization"*

Let us now turn to the absence of dynamism in the urban centers which is supposed to be the consequence of over-urbanization. The argument is that because this urban growth is abnormal, in the sense that it is not based on sufficient industrial development, the urban centers are not likely to be such dynamic centers of social and cultural change as, for example, they had been in Europe and other developed areas. This assumes that urbanization based on industrial development was mainly responsible for the social and cultural changes associated with urbanism or the urban way of life. This analysis is linked with the name of Wirth. Wirth's urban way of life, however, was a theoretical concept of a polar type that was hardly universal even in the Western industrialized countries. Wirth himself had become doubtful about his "ideal typical polar concept" towards the end of his life.<sup>14</sup>

Moreover, this line of thought regards the city as a key variable for explaining certain social phenomena and completely neglects the fact, emphasized by writers like Max Weber, that urbanization itself is a culture-bound phenomenon. Cities in various cultures diverge in some facets of their ecological and social structures. An interplay of these forces produces several types of urban communities even within the broad class of pre-industrial and industrial cities. Urbanization has its own universal structural, but they are not capable of influencing all social structures in the same way or with the same effectiveness. It is unrealistic to expect the same kind of social developments in underdeveloped countries as in the polar type of Western city. An infinite variety is possible.<sup>15</sup>

"...it is a highly debatable matter as to whether Western outlook, so characterized, is an antecedent or a consequent of industrialization and urbanization, or something of both, and also whether this outlook, or each element of it, really is an essential ingredient of economic development. It is conceiv-

in India," *Artha Vijnana* [Journal of the Gokhale Institute of Politics and Economics, Poona, India], June-September 1961.

<sup>14</sup> Passage quoted by Hauser from the posthumous collection of Wirth's writings in the UNESCO Seminar Report, *op. cit.*, p. 93.

<sup>15</sup> Cf. Gideon Sjoberg, *The Pre-Industrial City: Past and Present*, Glencoe: Free Press, 1960, p. 16.

able that the differences between Asian and Western outlook may produce somewhat different types of industrialization and urbanization or interpersonal and social relations arising therefrom. It is also conceivable that much of what has been written on the subject is the product of premature generalization based on limited observation of the Western experience."<sup>16</sup>

So far as the cultural role of the cities is concerned, there seems to be enough ground to believe that if viewed in relation to the rural areas that surround them, they do play a dynamic role.

"The primate cities of Asia are the most important centers of cultural change, especially in those fields which vitally affect economic development: advanced education, new forms of business organization, new administrative practices, and last but not least, new technologies find a fertile soil in them, their intermediate position between East and West, their contact with world markets of commodities and ideas, their land of many traditional bonds make them into eminently suitable vehicles for the introduction of new ideas and new techniques. If economic development is associated with modernization, the mediation of new, 'more modern' forms of social action through the primate cities of Asia is an indispensable part of this process."<sup>17</sup>

Rapid urbanization in underdeveloped countries is said to hamper economic development. Because of rapid urbanization the demand for provision of economic and social infra-structure investment increases much more rapidly there than in several other sectors. "This means that demand for less productive projects (in the immediate sense) will be made on the scarce capital resources of these economies in the early stages of their development."<sup>18</sup>

This would be so if the available capital resources in underdeveloped countries would be wholly or mainly devoted to more immediately productive investments. In regard to most of the underdeveloped areas, the infra-structure investments are necessary both in rural and urban sectors, and a large part of the

<sup>16</sup> *Ibid.*, p. 93.

<sup>17</sup> Bert F. Hoselitz, "Urbanization and Economic Growth in Asia," *Economic Development and Cultural Change*, Vol. VI, No. 1, October 1957.

<sup>18</sup> UNESCO Seminar Report, *op. cit.*, p. 8.

available capital resources has to go into them anyway. It is only a question of the relative share of the urban and the rural portions of the economy, with respect to the total investment to be devoted to it. In many ways, because of the earlier start in the cities in these countries, it is vastly more productive to make much of this kind of investment in the cities than in the less developed rural areas. They are likely to come to fruition much more quickly in the urban than in the rural sectors. Even apart from this, the greater advantages of quickly maturing as against slowly maturing capital investments in promoting economic growth are quite questionable. On the whole, therefore, the argument regarding the economic burden of rapid urbanization, as hampering economic growth in underdeveloped areas through the misallocation of scarce capital resources, is not impressive.

I have tried to show that the definition of over-urbanization that has emerged in the current discussion is unsatisfactory and vague; that the analysis of causes and consequences of over-urbanization developed so far is tenuous and oversimplified. Granted, however, that this is so, I am unwilling to end this discussion on a negative note by saying that this line of analysis, like that of overpopulation before it, is likely to end in a bog and to prove not very fruitful. I think that the time for drawing such a conclusion is not yet. The subject needs to be investigated further, perhaps by trying to go behind this analysis and to discern and then formulate in concrete, testable terms what it seeks to convey. I can imagine a few ideas that together or separately it struggles to express. The basic thought underlying the concept of over-urbanization seems to be some kind of undesirability of rapid urbanization in the underdeveloped countries. Perhaps it is felt that such a development is inimical to economic growth. In that case, it will have to be proved that in the absence of rapid urbanization, or at a slower pace of urbanization, these areas would have been able to progress more rapidly than they actually have so far. Without prejudging the issue it seems to me *prima facie* that this will be difficult to prove. Secondly, perhaps over-urbanization is felt to lead to the rise of new urban centers that are parasitical and also that the character of old urban centres becomes more parasitical because of over-urbanization. As a result, urbanization is not as creative as it should be. This will be difficult to verify but should be investigated, if this is what the present analysis seeks to convey. Thirdly, there seems to be implied some kind of norm of the tolerable density limits in rural and urban areas beyond which the

resulting social situation is somehow abnormal. If this is what is meant, then it will have to be much more concretely put, bearing in mind all the while how elastic in time and space such tolerable levels can be. There might be many other ideas behind the concept. Whatever they are, they should be formulated clearly and precisely before the concept and analysis of over-urbanization can be more solidly based and put to greater use.

# 19 The Economics of Urban Size<sup>†</sup>

William Alonso

## 1. INTRODUCTION

At least since Aristotle, men have wondered about the best size for cities. In the last decades developed and developing nations, capitalist and socialist have increasingly adopted more or less explicit policies on urbanization with special reference to city sizes. Most typically, these policies assume that the big cities of the nation are too big, and therefore try to disperse growth. Complementarily in recent years such dispersal policies, and policies addressed to distressed or backward regions, have recognized that these alternative centers must be of a certain minimum size, however ill-defined, in order to be viable. In its simplest sense the question of urban size consists of symmetric parts: how big is too big? and how big is big enough?

Theory and fact on these issues are scarce and poor, and they swim in an ocean of opinion, much of it highly emotional. For instance, recent American academic opinion, see Berry [6], Thompson [24], places the minimum size in the area of 250,000 to 500,000 inhabitants, which is about what the Soviet Union regards as the maximum tolerable size. The designation of "growth centers" has been awarded to settlements ranging in population from less than a handful of thousands to more than a million. Here we will present an aggregative economic approach to the theory of city size, and some empirical findings which suggest that even the largest cities have not yet reached excessive sizes from the point of view of growth and productivity. The discussion of least sizes will be more sketchy. Also, some observations and some data will be offered on the question of the size of a city in the context of its position in a system of cities.

## 2. THE DIFFICULTY OF DEFINING URBAN SIZE

Urban magnitude is no simple one-dimensional phenomenon. For instance an Asian metropolis of 5,000,000 inhabitants with a gross regional product per capita of \$100.00 has an economic magnitude equivalent to that of a modest American metropolis of 100,000, and is far smaller for income-elastic goods and services. Nor is the definition of population size unequivocal. Modern urban centers are surrounded by very large, diffuse zonal boundaries, within which there are marked

<sup>†</sup> The author wishes to thank the Centre for Environmental Studies, London, for their help. Research for this paper was supported by a grant from the Economic Development Administration, U.S. Department of Commerce.

Reprinted from *Papers of The Regional Science Association*, vol. 26 (1971), pp. 67-83, with permission of The Regional Science Association.

variations in the proportion of firms and people associated with that center, and in the intensity of the association. Thus, population does not constitute a conventional countable set, where people are unequivocally members or not. The situation is closer to that of "fuzzy sets," in which an element's membership in a set is a matter of degree. See Zadeh [28]. A number as a measure of population is thus gross oversimplification. The situation is further complicated in the frequent case of conurbations or megalopolitan areas, where the zonal boundaries of diverse centers overlap in complex patterns, and a person may be a member of two functional cities. In order to avoid the problem of the definition of boundaries (if not that of determining the degree of membership), some scholars have used density rather than size, but density is only a measure of local intensity of a ratio which misses the crucial aspect of the extension (or scale) of the system of interrelated elements. Certain villages of medieval origin have higher densities than giant metropolises or even their central cities. In general, we will ignore these problems, considering population as the basic magnitude and as a conventionally definable number, although some of the evidence to be cited is based on density measures; and towards the end of the paper I will consider some effects of adjacency of urban areas, although not the overlapping of their boundaries. Economic magnitude will be treated as an endogenous variable, basically a function of population.

### 3. THE MINIMUM COSTS APPROACH

Most approaches to city size have stressed the presumed diseconomies of urban scale, and have sought to establish that population at which costs per capita are least, regarding this as optimal. Although I shall argue that both the logic and the factual basis of this approach are faulty, it will be useful to review them briefly since they are so widely accepted by scholars and policy makers. Traditional studies of urban economies, going back to the turn of the century, have focused on how public costs vary as a function of population size using cross-section data. In general they have found them U-shaped, with the bottom of the curve occurring variously between 10,000 and 250,000 population. This literature is well known and has been supported by a few studies based on engineering calculations. Recent useful reviews are given by Cameron [8], Hirsch [12], Kain [14]. But the matter is not so simple, and these findings cannot be accepted at face value. Three principal difficulties may be mentioned. First, these measures of cost measure only inputs, and implicitly assume that outputs are constant. If the demand for public goods and services is at all income-elastic, cities with higher incomes would be spending more to get more, so that rising expenditures are not strong evidence of rising expensiveness. Indeed, the few studies based on multivariate techniques find no significant correlation of size and public cost after other variables are taken into account. See, for example, Schmandt and Stephens [22]. Second, the division between private and public costs is very much a matter of institutional convention. Most automobiles are private, but buses may be private or public. The production and distribution of electricity may be public or private. The sewer and waterpipes that vein a suburban residential district are public, but their exact



equivalents, running vertically within a large apartment house, are private. In brief, the category of public costs is neither well-defined nor stable. Third, many of the components of costs may not be real economic costs. For instance, suppose a teacher receives a higher salary in a large city because teachers there are unionized; the difference in his salary represents a transfer payment within the city rather than a true resource cost. It is unclear, in fact, how much of education is a production cost (training people) and how much of it is a form of consumption (educating people). Similarly, the treatment of land costs is ambiguous in the cost-benefit literature.

Contrast to the extensive literature on the costs of infrastructure and municipal operation, there is only a very slender literature on the variation in producers' costs with city size. No general study of the variation in producers' costs with city size appears to exist. A recent study by Morse [17] in an Indian context finds no substantial variation (and a possible small decline) in a range from very small cities to rather large ones. In the case of consumers' costs, it appears that these vary only slightly with urban size. See Alonso and Fajans [4]. The association is weak even for the housing and transportation components, which from theory might be expected to be strongly associated with urban size. The association disappears if other factors such as local climate and income are taken into account. Subjective estimates show, however, a sharp rise in the level of income that people think is needed for adequate levels of living in larger cities. See Gallup [11]. It is popular opinion, of course, that big cities are more expensive. It appears that one can live as cheaply in big cities as in small ones, but that the more varied opportunities of large cities raise expectations.

The most sophisticated explanation of the excessive growth of cities runs as follows: where costs are rising, a new industry (or inhabitant) makes its location decision on the prevailing (average) costs, including such factors as congestion and local taxes. However, since costs are rising, marginal costs are greater than average costs. Marginal costs are borne by the urban body as a whole, and the differences between average and marginal costs are the negative externalities. For instance, a plant considering locating in a large city will take into account existing (average) levels of congestion, but does not consider the increased congestion and travel costs that would be borne by the whole population as a result of its coming. By this argument, then, this divergence between private and social costs permits the city to grow beyond its best size.

It would be clearly impossible in real life to apply to each firm and citizen in order of arrival, such a differential tax, corresponding to the difference between marginal and average costs. Nonetheless, this view is reflected in the tax policies of many countries, such as Great Britain and France, which impose surtaxes on capital and/or labor in locations which are thought to be congested. Although such taxes operate on the average value for wages, they approximate a marginal approach for capital when they are levied on new investment. In many cases, of course, governments do not trust such uncertain subtleties of pricing, and public

policy manifests itself through direct command on industry and population, denying them some locations or even ordering them to others.

#### 4. AN AGGREGATE THEORY OF CITY SIZE

The argument of minimum costs is insufficient in its own terms. Such an objective is sensible only if output per capita is constant. But, in fact, it appears that output per capita is an increasing function of urban size. In that case, a more sensible objective of public policy would deal with the relation of outputs and inputs, rather than only with inputs. Before going on to spell out this simple point, we must return to definitional difficulties. It is in many cases most difficult to determine whether something is an input or an output. We have already raised this question with respect to education. But examples abound: for instance, are the expensive tie and suit that a businessman wears to his office production costs, or are they a form of consumption? Such problems, which had troubled early theorists of national income accounting, have been largely ignored by consensus in recent years. Yet they constitute a crucial area of ambiguity in the type of theory with which we are dealing, because they raise fundamental questions as to which human activities are instrumental and which are ends in themselves. For our purposes we will consider that urban output is the value of the total product of the urban area. Urban costs are harder to define, and would include quantity and price effects in the costs of infrastructure and municipal operation, in the costs of exogenous inputs other than human ones into the city's economic activity, and in private consumption. Thus, we regard the city as an aggregate productive unit.

Figure 1 shows a possible set of cost and product curves, and is akin to the traditional diagram of costs and revenues for the firm. The key difference is that the horizontal axis is in terms of population rather than of quantity produced. While the usual theory of the firm treats labor as an input whose price is exogenously determined, here labor is excluded from consideration in the construction of the cost curve, and the return to labor (in the broad sense of the total urban population) is the difference between the value of total output and total costs. Here I exclude from consideration colonial situations, in which part of this output is alienated by others elsewhere. The difference between costs and the value of output is thus available to the local population for investment or direct consumption, either directly from the city's production or through trade with others.

The average cost curve,  $AC$  in Figure 1, is shown as rising after a certain population level, both because this is generally believed to be the case (although the factual basis was questioned above) and because it weakens and therefore tests more sharply the argument being presented. The average product per capita curve,  $AP$ , is shown rising monotonically, partly to simplify the argument at this stage and partly because this is what most of the empirical evidence suggests. Later on we will consider the possibility of its turning down after some point.

If the general formulation is accepted, the rest is simple. The point of minimum per capita cost,  $P_c$ , is uninteresting. The point of maximum local contribution to national income occurs at  $P_e$ , where marginal cost,  $MC$ , is equal to marginal

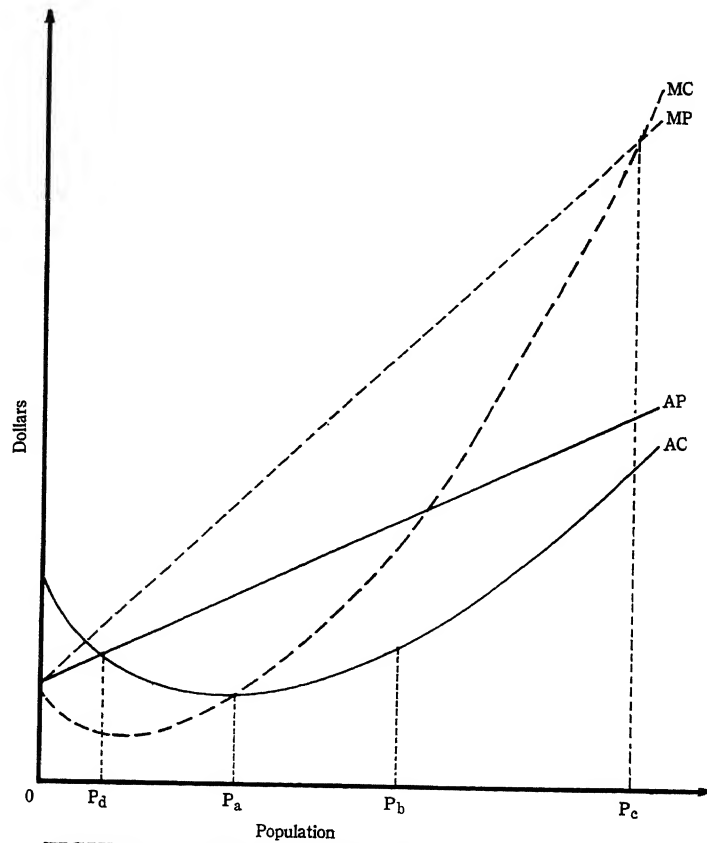


FIGURE 1. Urban Cost and Product Curves with City Size

product,  $MP$ . There is, of course, no need for this point to exist, in the sense that  $MP$  may remain higher than  $MC$  for the relevant range of population, but if it does exist, beyond this point further population costs more than it is worth. A national government interested exclusively in maximizing total product under conditions of labor surplus, would use such a population as its target. However, if there is not an unlimited supply of labor, the population size that would maximize national product would be smaller and would occur where the difference between  $MP$  and  $MC$  is equal to the opportunity costs similarly defined at alternative locations. From the point of view of the inhabitants of the city, however, a more sensible objective would be the maximization of the difference between average product,  $AP$ , and average cost,  $AC$ . This difference may be regarded as a per capita disposable income, and it will be maximized, of course, where the rate of increase of  $AP$  and  $AC$  are equal ( $dAP/dP = dAC/dP$ ).<sup>1</sup> Since average product is increas-

<sup>1</sup> This would have no meaningful equivalent in the context of the theory of the firm. It would amount to maximizing net revenue per unit product.

ing with population, this must occur at a size  $P$ , which is greater than that of minimum costs. In brief, the optimal population will differ according to whether a national or a local viewpoint is assumed, but in neither case will it coincide with the point of minimum costs.

From the point of view of policy, in trying to bring private costs and benefits into line with social costs and benefits, the same logic that leads minimum-cost theorists to argue for taxes on new arrivals based on the difference between social cost,  $MC$ , and private cost,  $AC$ , would argue here as well for subsidies based on the positive externalities produced by the new arrival. Thus, the net tax for subsidy would be based on a calculation of  $(MP-AP) - (MC-AC)$ , which may be positive or negative. Needless to say, these numbers are very elusive ones, and no operational tax could be based on such arithmetic.

There is no attempt here to develop a theory to explain the increase in production with population size. Pieces of such a theory exist in the literature and they stress economies of scale, advantages of specialization, agglomeration economies, probabilistic needs for lower relative reserves of inventories and labor, ease of communication, the richness of opportunities and the adaptability of large and complex systems, and so forth. See Thompson [25] and Alonso [1], [3]. Within the dishevelled theory of externalities, they suggest that positive external economies exceed the negative ones, principally by modifying the production functions of the component activities so that inputs are used more effectively. Without trying to formulate a general theory of causes, we pass to an examination of the empirical evidence.

## 5. EMPIRICAL EVIDENCE

In every country for which I have found evidence, local product per capita (or some index for it, such as income or wages) rises with urban size, and where comparable figures on cost are available, these rise far more slowly if at all. Although all of the data that one might wish for is not available for any single country, the overall pattern is clear. In some ways, the best figures available are those for the German Federal Republic and for Japan. West Germany offers estimates of Gross Community Product per capita, as shown in Table 1, which rises by 40 per cent from the smaller cities (20,000 to 50,000) to those above 500,000. Although public expenditures per capita rise by 44 per cent, the excess of product over expenditures rises by over 30 per cent from the smallest to the largest size class.

The data for Japan, developed by Mera [16] and shown in Table 2, are organized by density and by prefectures, with the densest areas corresponding to the largest cities. Income per capita (column 1) rises smoothly by nearly 80 per cent. While the curve of government expenditures is U-shaped (column 2), it is very shallow, partly because of the centralized nature of the Japanese system which, according to Mera, provides for quite a uniform level of public services throughout the nation. This would eliminate variations in these costs that arise from higher levels of demand generated by higher income, and leave only technological and price effects. Indeed, the Japanese relative variation in per capita government expendi-

TABLE 1. Gross Community Product and Public Expenditures  
Per Capita, German Federal Republic, 1964

| Population<br>Size Class | Gross<br>Community<br>Product per<br>Capita (DM) | Public Expend-<br>itures per<br>Capita (DM) | Public Expend-<br>itures as a<br>Percentage of<br>GCP | Excess of<br>GCP over<br>Public<br>Expenditures |
|--------------------------|--|---|---|---|
|                          | (1)  | (2)   | (2)/(1)   | (1)-(2)   |
| 20,000-50,000            | 5,400  | 513   | 9.5   | 4,900   |
| 50,000-100,000           | 6,000  | 627   | 10.5  | 5,400   |
| 100,000-200,000          | 5,900  | 603   | 10.2  | 5,300   |
| 200,000-500,000          | 6,400  | 738   | 10.4 <sup>a</sup>                                     | 6,400 <sup>a</sup>                              |
| 500,000-1,000,000        | 7,500  |   |   |   |
| 1,000,000 and over       | 7,500  |   |   |   |

<sup>a</sup> The arithmetic mean of the figures in column (1) was used for this composite class.

Source: based on figures from *Statistisches Jahrbuch Deutschen Gemeinden*, Band 52. Braunschweig: Waisenhaus Buchdruckerei, 1964.

TABLE 2. Japanese Population Density, Per Capita Income, Governmental  
Expenditures, Social Overhead Capital Stocks (SOCS) and all Govern-  
ment Investment Expenditure by Prefectures, in Thousands of Yen

| Population<br>Density<br>(persons per<br>Sq. Km.) | Number of<br>Prefectures | Mean per<br>Capita<br>Income<br>(1965) | Mean per<br>Capita<br>Government<br>Expenditure<br>(1965) | Mean per<br>Capita<br>SOCS<br>(1963) | Mean per<br>Capita All<br>Government<br>Investment<br>Expenditure<br>(1966) |
|---|--------------------------|--|---|--------------------------------------|---|
|   |                          | (1)                                    | (2)   | (3)                                  | (4)   |
| Less than 200                                     | 17                       | 188.0                                  | 50.8  | 268.5                                | 32.8  |
| 200-300   | 15                       | 197.6                                  | 46.7  | 244.8                                | 27.3  |
| 300-600   | 8                        | 209.1                                  | 43.2  | 206.4                                | 27.6  |
| 600-1,000   | 3                        | 228.0                                  | 41.9  | 205.3                                | 27.5  |
| 1,000-3,000                                       | 1                        | 280.0                                  | 53.7  | 178.4                                | 40.5  |
| 3,000 and over                                    | 2                        | 340.0                                  | 56.1  | 188.0                                | 39.4  |

Note: 1960 population in all cases.

Source: adapted from Mera [16].

ture from highest to lowest is only about half that of the American or the German equivalent.

Perhaps the most surprising element in the data is the sharp decline with increasing density in Social Overhead Capital Stocks (SOCS) per capita (column 3). This runs counter to common belief that the bigger the city the more infrastructure per capita is needed, and may be the result of such effects as the greater linear quantities of roads per capita necessary in low density areas. The lowest SOCS per capita are for the penultimate density class, but this class consists of a single prefecture adjacent to Tokyo, and presumably shares some of the advantages of Tokyo's size. Yearly government investment expenditures (column 4), which may be interpreted as marginal SOCS, are also U-shaped, declining at first and

TABLE 3. Ratio of Actual to "Expected" Hourly Earnings by City Size, United States, 1959

|  | White Males | White Females | Non-White Males | Non-White Females |
|--|-------------|---------------|-----------------|-------------------|
| <i>Rural</i>                                   | 0.83        | 0.83          | 0.78            | 0.76              |
| <i>Urban Places</i>                            |             |               |                 |                   |
| Under 10,000                                   | 0.84        | 0.84          | 0.75            | 0.63              |
| 10,000-99,999                                  | 0.91        | 0.88          | 0.76            | 0.78              |
| <i>Standard Metropolitan Statistical Areas</i> |             |               |                 |                   |
| Under 250,000                                  | 0.97        | 0.94          | 0.84            | 0.76              |
| 250,000-499,999                                | 0.97        | 0.96          | 1.04            | 0.90              |
| 500,000-999,999                                | 1.02        | 1.03          | 1.07            | 1.00              |
| 1,000,000 and over                             | 1.12        | 1.13          | 1.10            | 1.19              |

Note: "Expected" hourly earnings were calculated for each sex and color by multiplying the national average hourly earnings of each color and sex by age and education by the annual hours worked by members of that cell in the region, summary in each case across all such cells in the region and dividing by the total man-hours of the region.

Source: adapted from Fuchs [10].

rising for the densest two classes. However, such investment may be expected to be more a function of the rate of population growth than of the actual density, and the most urbanized areas are receiving a greater share of immigrants.

Turning now to American data, Fuchs [10] found a steady rise of about 40 per cent in hourly earnings with urban size after discounting the effects of labor composition by color, age, sex, and education. See Table 3. If, by traditional economic theory, we equate wages to the marginal productivity of labor, the conclusion is that, for a given type and quality of labor, the rise in wages indicates that productivity rises with urban size. Comparable data, standardized by industrial composition rather than by the characteristics of workers, is shown in Table 4. Although allowance must be made for the difference in the range and disaggregation of the population classes, the results are fairly similar to those of Fuchs. The Payroll per Employee Index (column 1) rises with urban size, although it flattens over the last two classes. Value Added per Employee (column 2) rises a bit more strongly. Taking the wage component out of value added, the remainder of value added (column 3) shows a very strong uninterrupted rise which is the more remarkable in view of the decline of new capital per employee (column 4). Information on capital stocks was not available. Thus, the rise in value added per worker can be attributed neither to rising wages nor to more massive use of capital, and may be attributed to the positive externalities of urban size. We shall return to this below.

Many studies over the years have established that in the United States, for data grouped by city size, there is a strong and steady rise in income, of about 30 per cent depending on the size of classes and income definition used. In keeping with sociologists who have looked into various correlates of urban size, Schnore [23]

TABLE 4. Indices of Payroll, Value Added, Value Added minus Payroll, and New Capital, by Employee, Standardized by 2-Digit S.I.C., 67 Largest Standard Metropolitan Statistical Areas, United States, 1963

| Population          | Payroll per Employee Index<br>(1) | Value Added per Employee Index<br>(2) | Value Added minus Payroll per Employee Index<br>(3) | New Capital per Employee Index<br>(4) |
|---------------------|-----------------------------------|---------------------------------------|---|---------------------------------------|
| 250,000-500,000     | 0.994                             | 0.970                                 | 0.943   | 0.967                                 |
| 500,000-1,000,000   | 1.029                             | 1.047                                 | 1.099   | 1.028                                 |
| 1,000,000-5,000,000 | 1.061                             | 1.046                                 | 1.029   | 0.925                                 |
| 5,000,000 and over  | 1.058                             | 1.053                                 | 1.129   | 0.784                                 |

Source: based on data in Douglas [9].

has concluded that, "Of all the differences among communities of different size revealed in this study, perhaps the most striking is the pronounced direct relationship between size of place and income." However, since group data tends to subsume intra-class variance, I have experimented with regressions of income on population. See Alonso and Fajans [4]. In a simple logarithmic regression (not reported in that paper) metropolitan population accounted for 42 per cent of the variance in mean per capita incomes. This is indeed remarkable if one considers the great diversity among cities of climate, resource endowment and history. The relation is only slightly diminished when metropolitan incomes are deflated by local cost of living in the smaller set of metropolitan areas for which the information is available. While local government expenditures per capita rise from \$120 at populations of 15,000 to 50,000 to \$200 in counties of more than 1,000,000 population, see Schmandt and Stephens [22], this rise, which represents \$250 per family, is made insignificant by a rise of \$1500 in income over the same range.

Higher incomes are found in bigger places in developing countries and in socialist countries as well, although there are significant understandable exceptions. Thus, the per capita income of Rome is lower than that of the great industrial cities of the Italian north, and in some developing countries there are some steel or oil cities where incomes are exceptionally high. Similarly, Birmingham, Alabama, a steel town, has a higher payroll per employee than does New York. Particular circumstances are always at work, and here we are speaking of general patterns which are useful for policy analysis, rather than for particular project evaluation.

Similar questions are being raised in the Soviet Union. Perevedentsev [19] observes that the extensive Soviet literature on city size, while speaking of "excessive growth," "excessive concentration," and the like, offers no criteria for determining what is excessive. He notes the continued growth of larger cities, in spite of strong direct controls on personal and industrial location to stop this growth, while the productivity of labor is 38 per cent higher and the return on assets is more than twice as high in cities of more than 1,000,000 than in cities of between 100,000 and 200,000. In contrast to these facts, he observes that, "according to the prevailing views of our city planning, cities with a population of 50,000 to 200,000 are considered



optimal, and those with up to 400,000 are permissible." In a reply, Khorev [15] speaks of the particularity of cities, especially in their industrial composition, of the higher costs of infrastructure in large cities, and of the value of the additional time spent in daily commuting. He shifts the question of optimal size to one of the particular circumstances of a city and to the hierarchy of cities, but does not offer operational guidelines or systematic documentation.

Since we conclude that it appears that the biggest cities are not too big from the viewpoint of economic efficiency, it may be asked whether the higher average incomes of bigger cities do not mask sharper inequalities among their citizens, so that efficiency is gained at the cost of equity. This does not seem to be the case, at least for the United States. On the contrary, some recent studies, such as those by Ornati [18] and Burns [7], indicate that there is less poverty and a more equal distribution of incomes in big cities than in smaller ones. In the pages below, I will touch on the relation of urban sizes and interregional disparities.

## 6. A POSSIBLE DOWNTURN IN THE PRODUCT CURVE

A question of particular interest, even if the general rise of income with size is granted, is whether there is a turndown in the relation after some size. I have tried very hard to test for this with income data for the United States, and in some cases equations using the equivalent of a polynomial of population (i.e., population and its natural logarithm) do yield such a turndown, somewhere between 3,500,000 and 9,000,000 population. However, the standard error of estimate as to where the inflection point occurs is extremely high, and not even New York can be excluded with any certainty from the rising trend, even taking into account differences in cost of living. See Alonso and Fajans [4]. Nonetheless, the possibility of such a downturn is of extreme importance, for it would suggest that those cities were in some sense too big for that society at that time. Therefore, the question must be asked: assuming that a turndown does exist, however weak its indication, does this mean that cities larger than the inflection point are too large? Three reasons suggest themselves why this might not be the case:

- 1) The development of very dispersed patterns of suburbanization far beyond the territorial definition of the standard metropolitan statistical area, sometimes called exurban, is commonly observed in the United States and has been recently strikingly documented by Berry [6]. These exurbanites typically enjoy greater incomes than the mean of the SMSA population, and are not included in the statistics for the SMSA. If the relative importance of this phenomenon is larger for larger areas, as anecdotal evidence suggests, the mean income of the larger areas would be understated, and the downturn would be more apparent than real. However, it is no easy matter to test for this, for this dispersed metropolitan population is only a fraction of the population resident in the surroundings of the SMSA, and because the largest metropolitan areas are set in a context of other nearby metropolitan areas whose fuzzy zonal boundaries overlap.

- 2) On the other hand, the decline might be real and yet not signify that these largest cities are too large. They have traditionally performed a port of entry



function for foreign and agricultural populations which have low productivity and earning power during the years and even generations of their acculturation. As urbanizing centers they perform a systemic function for the nation as a whole, and other cities benefit in time as the acculturated citizens disperse to other centers. Inclusion of a non-white variable, as a rough test of this hypothesis, weakens or eliminates the turning down of the curve.

3) Lower incomes and lower wages in the largest cities could be consistent with their efficient size in a nation with a high rate of innovation. It has been argued by several authors in recent years that the largest cities serve as seedbeds for new economic activity at a time when its technology is in rapid flux, the nature and extent of its demand unknown, and its institutional structure unformed. See Hirschman [13], Vernon [26], Alonso [1], [3] and Thompson [24], [25]. During these formative years such new activities depend on an environment rich in external support, where there is available direct access to customers, financing, shipping, jobbing, specialized labor, rumors, ideas, and a thousand other things. If the enterprise grows and succeeds, its management, marketing, and production become routinized into well-defined roles, its technology and demand patterns stabilize so that it becomes possible to standardize and routinize, and hence to substitute capital for labor. With size and predictability, many of the externalities can be internalized. At this point, industries may migrate to provincial centers (often through branch plants) seeking advantages of transportation costs, of cheaper and plentiful labor, of particular linkages with suppliers, or of economies in other inputs or distributional advantages. This pattern might be expected to lead to lower earnings in the centers of innovation because the fluidity of infant industries logically leads to low capitalization, and from this to a lower marginal productivity of labor, and therefore lower wages. This evolution has often been described and documented anecdotally. The most general historical documentation is that by Pred [20]. The cross-sectional data of Table 4 supports this view. The rise in Payroll per Employee suggests that the lower wages arising from lower capital inputs are to some degree compensated by the greater productivity of the larger cities. But the Value Added minus Payroll per Employee column, which discounts wages and rises sharply with urban size, stands in sharp contrast with the overall decline in the New Capital per Employee Index column. In a country such as the United States, where capital markets are relatively well integrated, it is not credible that there should be such sharp differences in returns to capital at different locations as these two columns would indicate. Rather, after allowance of returns to capital, it would appear that there are strong returns to a factor that is often neglected in this type of analysis because it is hard to measure: this is the return to entrepreneurship, or entrepreneurial rent and wages. Within this I would include the return to institutional forms when these constitute a local resource. Thus the data in Table 4 are perfectly consistent with this theory although, of course, they do not prove it. It would be possible to have slightly lower income or wages for the largest cities, consistently with overall national efficiency on the basis of the dynamic and innovative systemic function of these centers.

To summarize: there are some weak indications of a downturn in the product curve at the largest urban sizes. However, even should the downturn be real, this would not be inconsistent with the efficiency of those larger sizes in a hierarchical system of cities.

## 7. THE MINIMUM SIZE OF CITIES

Many countries are today engaged in policies for the promotion of economic growth in distressed regions and in frontier regions. Although there seems to be a general agreement in the use of growth centers to concentrate development efforts, the theoretical and factual bases are even weaker for small cities than for large. Economic base, input-output, and other multiplier theories are insensitive to scale effects and to positive or negative externalities, and most other theory seems to hover somewhere between intuition and poetry, with reference to thresholds, will to develop, industrial vocation, propulsive industries to serve as the engine for development, and the like.

This uncertainty as to size is illustrated by a range of three orders of magnitude in the population of growth centers from one country to another. Not surprisingly, empirical studies are extremely scarce since there is lacking a theoretical foundation on which to base factual questions. A few recent studies in the United States have hazarded guesses about a minimum population size for self-sustaining growth. Berry [6], on the basis of a break in the rank-size relation, suggests the quarter million mark. Thompson [25], on the basis of recent growth statistics, places the level somewhat higher. Alonso and Medrich [5], also on the basis of historic growth patterns, identify spontaneous growth centers throughout the size hierarchy. It seems unlikely that there would be a well-defined threshold. Small economies may be expected to be much more particularistic than large ones, where something similar to a law of large numbers might be expected to operate. In a small economy, where the elements are few and the connectivity of linkages therefore relatively low, particular events and circumstances may be expected to play a larger role. The fortunes of a particular corporation, or the sudden economic relevance of a local resource or locational advantage can produce strong rates of growth or decline in a small economy, while in a large one there is always a very strong probability that compensating forces will dampen fluctuations. Although there are strong overall patterns in the rates of growth by urban size, the smaller centers exhibit a far greater variation in their rates.

The greater particularity of smaller centers and the resulting greater variance in their economic and demographic rates, make an aggregative approach such as that presented below of relative little usefulness for the consideration of any one particular small center. Such an aggregative approach is meant to shed light on general policy rather than on a particular program of project evaluation. If, as suggested, the firm perceives its costs and revenues as average costs and average product, examination of the left side of Figure 1 shows clearly why a firm might find disadvantageous a location in a small city while, at the same time, such a location might be socially valuable. To simplify the argument, consider for the moment

that wages are included in the costs. At populations smaller than  $P_d$ , a firm would, in fact, be losing money; a location would not become attractive until the difference between  $AP$  and  $AC$  became comparable to the firm's opportunity costs at other locations. The threshold of spontaneous self-sustaining growth would occur at that point. Nonetheless, if productivity increases with size, marginal product will be higher from the social point of view than average product even at small sizes; and if costs are declining, social marginal costs will be lower than the costs as perceived by the firm. Therefore, it might make excellent sense to subsidize such a firm, either directly or through infrastructure, to induce it to locate in the growth center. This is particularly so if society has a longer range view and is considering a substantial increment in size from the location of several firms, while the rise from the location of one firm would be too small. The great difference is that in large cities the external benefits foregone by the firm (the difference between marginal and average product) are compensated by the implicit subsidy of its paying average costs, which are lower than marginal costs (we are assuming that costs are rising). In the smaller city, the differences between average and marginal rates of both cost and product operate to the disadvantage of the firm.

Such a subsidy to small cities would be justified from the point of view of national efficiency only if the opportunity costs, again defined as the difference between the marginal rates, were lower in the large cities than in the proposed growth center. Our empirical evidence suggests that this is seldom the case. Only in an economy with a labor surplus, with capital costs included within the cost curves, would there be no opportunity costs to the location of population, and would a general policy of development of smaller centers be consistent with a national goal of efficiency. But such surplus populations are common only in developing countries, and in these the scarcity of innovational, managerial, and institutional resources make unrealistic such an urbanization policy.

The development of smaller centers in distressed areas is more likely to find this justification in equity or distributional objectives. A small center might be aided to the point of self-sustaining growth for the purpose of improving local levels of welfare, but this will typically be done at some cost to national product. For instance, Mera on the basis of similar reasoning estimates that the equalization of per capita incomes by prefectures in Japan would reduce national income by 15 to 30 per cent. While the objective of equalization of interregional levels of welfare is a widely accepted fact, much national policy (such as that of the United States) asks simultaneously for conflicting goals: the minimizing of disparities together with the maximization of output. A similar logical contradiction occurs frequently in attacks on large cities. They are charged at one and the same time with: a) being grossly inefficient, in that increases in population cost more than they produce; and b) increasing interregional disparities in levels of welfare through the disproportionate concentration of economic growth. Both of these can only be true under very unlikely circumstances, see Alonso [2], because if big cities were very inefficient, their incomes (or the difference between product and costs) would decrease with increasing size, thus reducing inequalities. However, the simple cross-sectional

analysis of this paper does not pretend to cover the complex dynamics of the long-run spatial aspects of development, in which growth, concentration, and inequality appear in varying relationships at different stages. See Hirschman [13], Alonso [1], [3], Williamson [27].

#### 8. CITY SIZE AND ACCESS TO OTHER CITIES

In a fundamental sense, the population of a city may be regarded as an index of the number of opportunities available to a person or enterprise in that city. Just as productivity might be expected to rise with the wealth of opportunities in a city, it might also be expected to rise with the nearness to other opportunities in other cities. To test this, the 1959 mean per capita income,  $y$ , of 211 American metropolises was correlated with population,  $P$ , and with their income potential. Income potential,  $V$ , for each metropolis was calculated as the sum over all other metropolises of their population times their mean income, divided by their distance in kilometers from the metropolis in question. Potential at a certain location may be interpreted as the accessibility to the constellation of urban opportunities available to a person or a firm at that location. The resulting equation was obtained by regression:

$$y = e^{5.01} P^{0.0681} V^{0.0866}$$

$$R^2 = 0.26$$

$t$  values of exponents: 33.3, 6.3, 4.9.

Again, considering that we are disregarding local resources, social, economic, and political history, locational advantages, climate, and many other factors, it is remarkable that these two simple and very gross independent variables should account for better than one fourth of the variance in incomes.

The effect of income potential in the equation is to raise the intercept of the curve of productivity with population for cities accessible to the urban population as a whole. If we assume that urban costs are a function of size but not of location, it follows that growth centers or new towns in distant regions must be bigger to arrive at self-sustaining growth than centers in more accessible ones. Or, in program terms, that the extent of subsidies must be greater for growth centers in low potential areas. Conversely, this finding is also significant to an understanding of the megalopolitan phenomenon. Again, assuming that costs are a function of urban size but that productivity is a function of size and potential, if the diseconomies of size are strong, it would follow that the pattern of development would tend to favor small metropolitan areas in the vicinity of large ones, or clusters of strongly interconnected medium-size cities. These smaller metropolises would be in the position of having their cake and eating it, for they would enjoy the advantages of agglomeration without the disadvantages of size. The pattern then would be one of megalopolitan development, which as an adaptation to ever-increasing populations extends the shift from the mononucleated large city of the nineteenth century, to the polynucleated metropolis of mid-twentieth century, to urbanization in the

form of a constellation of metropolitan areas. This development is observable in most developed countries, where many of the fastest growing metropolitan areas are the smaller ones within the megalopolitan complexes. See Alonso and Medrich [5].

It is a pity that the term *megalopolis* has acquired negative connotations in journalistic usage. There is a popular image of a teeming anthill, of urban areas running into each other and choking up open space. This is misleading. For instance, the first-diagnosed megalopolis is the Mid-Atlantic one in the United States. By a broad areal definition only 6.0 per cent of its territory was built up in 1960, 12 per cent by a more restricted definition, and only 21 per cent of the territory within its component Standard Metropolitan Statistical Areas. See Regional Plan Association [21]. These statistics will be confirmed by anyone who has flown over the area and seen the vast open spaces within which the urbanized areas huddle. If there is a propensity for growth by the smaller units of megalopolis, there is no lack of space for that of growth.

I have argued in this paper that it is a mistake to emphasize the variation in urban costs with urban size and to neglect variations in productivity. Similarly, it is misleading to consider only size, which is a measure of immediate opportunities, while neglecting the broader context of opportunities in other cities. Big and small must be qualified in their setting: whereas it may be quite good to be smaller in a dense setting, it may be necessary to be quite big in an isolated one. Policies of small and far, which are not uncommon, perhaps should be small and near, and big and far.

#### 9. PROBLEMS IN THE INTERPRETATION OF THE CURVES

It would be well to end this discussion with an upbeat note, but it must be concluded with further cavils of interpretation and definition. Curves such as those in Figure 1 will vary from city to city, from society to society, and within a society with its evolution through time. Even after the improvement of the contextual variable of income potential, the population variable accounts for only part of the variation in local incomes. Further, while the empirical data shows substantial consistency across the arrays of cities of particular nations, it is based primarily on cross-sectional data. One might ask what is the meaning for a city of 20,000 of curves extending into the tens of millions; or, conversely, what is the meaning for a great metropolis like New York or Tokyo of the range of the curve at the 20,000 level.

It must be granted that the relation of size to time is far from clear in these arguments. Surely it is not the same to grow tenfold in a decade as in a century, and movements along the population axis in Figure 1 would take place in real time and in particular circumstances. One might ask, further, whether paths of expansion coincide with paths of retrenchment, whether a city grown too big can regain paradise by shedding its excess population. In brief, the analysis admittedly suffers from the common limitations of the application of static theory to a dynamic process. Most particularly, there is no suggestion here that a static general equilibrium

approach can serve to model a nation's system of cities because the movements of people, capital, ideas and institutional forms are slow and evolutionary, and while they change, other things are changing. Issues such as these have more to do with the dynamics of fluids than with the mechanics of solids.

## REFERENCES

- [1] Alonso, W. "Industrial Location and Regional Policy," Working Paper No. 74, Center for Planning and Development Research, University of California, Berkeley, 1968.
- [2] ———. "Equity and Its Relation to Efficiency in Urbanization," in J. F. Kain (Ed.), *Essays in Regional Economics*. Cambridge: Harvard University Press, forthcoming.
- [3] ———. "Urban and Regional Imbalances," *Economic Development and Cultural Change* Vol. 17 (October 1968), pp. 1-14.
- [4] ——— and M. Fajans. "Cost of Living and Income by Urban Size," Working Paper No. 128, Center for Planning and Development Research, University of California, Berkeley, 1970.
- [5] ——— and E. Medrich. "Spontaneous Growth Centers in Twentieth Century American Urbanization," in N. Hansen (Ed.), *Growth Centers and Regional Development*. New York: The Free Press, forthcoming.
- [6] Berry, B. J. L. and E. Neils. "Location, Size, and Shape of Cities as Influenced by Environmental Factors: The Urban Environment Writ Large," in H. Perloff (Ed.), *The Quality of the Urban Environment*. Baltimore: The Johns Hopkins Press, 1969.
- [7] Burns, L. S. "The Intra-Urban Income Distribution," Paper presented at the Annual Conference, British Section of the Regional Science Association, London, 1968.
- [8] Cameron, G. C. "Growth Areas, Growth Centres, and Regional Conversion," *Scottish Journal of Political Economy*, Vol. 17 (February 1970), pp. 19-38.
- [9] Douglas, R. "Selected Indices of Industrial Characteristics for U.S. Standard Metropolitan Statistical Areas, 1963," Discussion Paper No. 20, Regional Science Research Institute, Philadelphia, 1967.
- [10] Fuchs, V. R. *Differentials in Hourly Earnings by Regional and City Size, 1959*. New York: National Bureau of Economic Research, Occasional Paper No. 101, 1967.
- [11] Gallup, G. *Typical U.S. Family of Four Needs \$120 per Week to Make Ends Meet*. Princeton, N. J.: American Institute of Public Opinion, 1970.
- [12] Hirsch, W. Z. "The Supply of Urban Public Services," in H. Perloff and L. Wingo (Eds.), *Issues in Urban Economics*. Baltimore: The Johns Hopkins Press, 1968.
- [13] Hirschman, A. O. *The Strategy of Economic Development*. New Haven: Yale University Press, 1958.
- [14] Kain, J. F. "Urban Form and the Cost of Urban Services," Discussion Paper No. 6, Program on Regional and Urban Economics, Harvard University, 1966.
- [15] Khorev, B. "What Kind of City is Needed," *Literaturnaya gazeta*, No. 14, 1969, as reported in *The Current Digest of the Soviet Press*, Vol. 21 (April 23, 1969), p. 14 and p. 27.
- [16] Mera, K. "On the Concentration of Urbanization and Economic Efficiency," Economics Department Working Paper No. 74, International Bank for Reconstruction and Development, Washington, D.C., 1970.
- [17] Morse, R. et al. "Costs of Urban Infrastructure as Related to City Size in Developing Countries: India Case Study," Stanford Research Institute, Menlo Park, California, 1968.
- [18] Ornati, O. A. "Poverty in Cities," in H. Perloff and L. Wingo (Eds.), *Issues in Urban Economics*. Baltimore: The Johns Hopkins Press, 1968.
- [19] Perevedentsev, V. "Cities and Years," *Literaturnaya gazeta*, No. 9, 1969, as reported in *The Current Digest of the Soviet Press*, Vol. 21 (March 19, 1969), pp. 6-9.
- [20] Pred, A. *The Spatial Dynamics of U.S. Urban and Industrial Growth, 1814-1800*. Cambridge: M.I.T. Press, 1966.

- [21] Regional Plan Association (New York). "The Atlantic Urban Seaboard: Development Issues and Strategies," *Regional Plan News*, No. 90 (September 1969), pp. 1-33.
- [22] Schmandt, H. J. and G. R. Stephens. "Local Government Expenditure Patterns in the U.S.," *Land Economics*, Vol. 39 (November 1963), pp. 395-406.
- [23] Schnore, L. "Some Correlates of Urban Size: A Replication," *American Journal of Sociology*, Vol. 69 (September 1963), pp. 185-193.
- [24] Thompson, W. R. *A Preface to Urban Economics*. Baltimore: The Johns Hopkins Press, 1965.
- [25] ———. "Internal and External Factors in the Development of Urban Economies," in H. Perloff and L. Wingo (Eds.), *Issues in Urban Economics*. Baltimore: The Johns Hopkins Press, 1968.
- [26] Vernon, R. *Metropolis 1985*. New York: Doubleday 1963.
- [27] Williamson, J. "Regional Inequality and the Process of National Development," *Economic Development and Cultural Change*, Vol. 13 (July 1965), Part 2.
- [28] Zadeh, L. A. "Toward a Theory of Fuzzy Systems," Report No. ERL-69-2, Electronics Research Laboratory, University of California, Berkeley, 1969.



## 20 The Urban Poor: Disruption or Political Integration in Third World Cities?\*

Joan Nelson

MOST of the nations of Africa and Asia remain predominantly rural and agricultural. However, more than half the people in most Latin American countries are no longer rural, and a fifth to a third live in cities of 100,000 or more. In Asia and North Africa, Lebanon, the U.A.R., and the Philippines are also substantially urbanized, and Morocco, Syria, Turkey, South Korea, and Taiwan are not far behind. Moreover, virtually everywhere in the developing world, regardless of the extent of urbanization already achieved, cities are growing at rates of 5 to 8 percent annually. That is, they are doubling their populations every ten to fifteen years.<sup>1</sup>

Such rapid growth has far-reaching social, economic, and political repercussions. Rural areas are steadily losing many of their more ambitious young adults. A large and growing share of the cities' populations are newcomers to urban life. Exploding physical size and pressure on urban services impinge on long-established urban groups. All of

\* I am grateful to the Center for International Affairs of Harvard University and to the Council on Foreign Relations for their financial support and intellectual stimulation during the period when this study was developed. Special thanks for thoughtful and constructive criticism of earlier drafts are due to Peter Bell, Donald Horowitz, and Harold Wilensky.

<sup>1</sup> In Brazil and Mexico between 1940 and 1950, population in cities of 100,000 or more grew at average annual rates of 5 and 6.7% respectively. During the 1950's, Santo Domingo grew 7.3% yearly; Panama City expanded at a rate of 7.9%. (United Nations *Compendium of Social Statistics*, 1963, Series K, No. 2, Table 7.) In the 1950's and early 1960's, Bogotá's population rose an average of 6.8% a year; Cali's increased at 6.3%. Paul T. Schultz, *Population and Labor Force Projections for Colombia, 1964-1974*, mimeo. (Santa Monica, California, RAND, July 10, 1967), 12. Between 1941 and 1959 Caracas averaged a 7.4% annual growth. Bruce Herrick, *Urban Migration and Economic Development in Chile* (Cambridge, Mass. 1965), 31. In some other parts of Asia and the Near East, rapid urban growth rivals that of Latin America. Korean cities have been growing rapidly since the 1950's: Seoul added 6.6% more people each year from 1960 to 1966. Turkey's population centers of 100,000 or more grew 6.7% a year from 1955 to 1960; Ankara averaged 6.8% annually from 1960 to 1965. (Estimated from figures in the United Nations *Demographic Yearbook*, 1962, 1963, 1967.) In South Asia, urban growth rates are generally lower. Delhi grew 5% a year from 1951 to 1961, but greater Bombay expanded at an annual rate of 3.9% during that period, and Calcutta's rate was 1.9%, reflecting in part the immense size already reached by these two giants. Kingsley Davis in Roy Turner, ed., *India's Urban Future* (Berkeley 1962), 10.

Reprinted from *World Politics*, vol. 22, no. 3 (April 1970), pp. 393-414, with permission of the author and Princeton University Press. © 1970 by Princeton University Press.



these changes have political effects. Yet to date there has been remarkably little analysis of the political effects of rapid urban growth.

This paper focuses on only one set of effects: the immediate and longer-run political role of migrants to the larger cities, and more generally of the urban poor. It argues that, contrary to widespread speculation by both foreign observers and elites in the countries concerned, neither new migrants nor the urban poor are likely to play a direct destabilizing role. Indeed, historical experience in industrialized nations suggests that the urban poor are not likely to play a significant political role at all. However, there are major demographic, economic, and political contrasts between today's developing nations and Europe and North America in the late nineteenth and early twentieth centuries. The differences in context make it possible that the urban poor of Latin America, North Africa, and Asia may come to take a more active part in municipal and national politics than did their historical counterparts. The paper concludes with some suggestions regarding possible patterns of political participation by the urban poor.

#### I. URBAN GROWTH AND POLITICAL INSTABILITY<sup>2</sup>

Many of the migrants swelling the cities have little or no education and few skills. Moreover, in most developing countries rates of unemployment and particularly underemployment are high and rising. Programs to improve conditions in rural areas are often urged as a means to stem the movement to the cities. But the very processes that promote rural modernization and integration—education, mass media, transportation—also spur the exodus in search of a better life. Labor-intensive techniques in industry are sometimes proposed as a means of absorbing laborers once they reach the cities. But such techniques often imply high unit costs of production, and are unlikely to be adopted on a large scale. In short, the cities face a steady rise in the proportion of their populations that is unskilled, semi-employed, and abysmally poor—marginal to the city's economy and social organization.<sup>3</sup>

<sup>2</sup> For a more detailed discussion of the points in this section, see Joan M. Nelson, *Migrants, Urban Poverty, and Instability in Developing Nations*, Harvard Center for International Affairs, Occasional Paper No. 22 (Cambridge, Mass. 1969).

<sup>3</sup> Open unemployment rates of 5 to 10% appear repeatedly in surveys of major Latin American and Indian cities. Rates are higher among the unskilled, among young men seeking their first jobs, and in larger cities. Fred Dziadek, *Unemployment in the Less Developed Countries*, AID Discussion Paper No. 16 (Washington, June 1967), Appendix A. Urban underemployment must greatly exceed unemployment, but it is extremely difficult to measure. A proxy indicator of the extent of underemployment is productivity in the tertiary or service sector, which drops as the sector is

From the viewpoints of welfare and economic development, the growth of a large marginal urban population is not necessarily detrimental. Despite insecure jobs and squalid living conditions, most migrants feel that they are better off than they were before coming to the city. The movement to the cities facilitates wider distribution of certain services, particularly schools and clinics. Concern on grounds of welfare therefore is probably misplaced. Rural-to-urban migration also removes redundant labor from agriculture. While productivity in the overloaded service sector is falling in many developing nations, it is still substantially higher than productivity in agriculture. Though separate data are not available on productivity in intermittent day labor or construction jobs and in low-paid segments of the service sector (vending, loading and carrying, domestic help, petty personal services), these ways of scraping by are probably not less productive than marginal production in agriculture. Moreover, residence in the city exposes the migrant to modernizing influences, and improves his opportunity to acquire skills, however modest. The major economic disadvantage is the diversion of public funds and energies from more directly and immediately productive uses that is likely to result from pressure for expanded municipal services.

The strongest concern about rapid urban growth and urban underemployment is political rather than economic. A lengthening list of articles asserts that chaos and revolution lurk in the *favelas* of Rio and the alleys of Calcutta.<sup>4</sup> Dire predictions about the political effects of rapid urban growth take two main forms.<sup>5</sup> Some of the prophets focus on the migrants. Uprooted, isolated, disappointed, and frustrated, they are viewed as tinder for any demagogic or extremist spark. Other theorists argue just the opposite: new migrants are politically passive.

---

swollen by peddling, domestic service, and other marginal occupations. It has been calculated that in Latin America between 1950 and 1965, while productivity in agriculture grew at 1.8% a year and industry, mining, and utilities at 2.5% annually, productivity in the service sectors fell, suggesting "a level of underemployment equivalent to 10% of the national labor force." Since services absorb much more urban than rural labor, the implied underemployment rate in the cities would far exceed ten percent. Hollis B. Chenery, *Toward a More Effective Alliance for Progress*, AID Discussion Paper No. 13 (Washington 1967), 12.

<sup>4</sup> See, for example, Barbara Ward, "The Uses of Prosperity," *Saturday Review*, August 29, 1964, 191-92; Franz Fanon, *The Wretched of the Earth* (London 1965), 103.

<sup>5</sup> Peter Lupsha's interesting article, "On Theories of Urban Violence," presented at the American Political Science Association meetings in 1968, lists many more theories of the causes of urban violence, including "conspiracy," "riff-raff," "teen-age rebellion," and "police brutality." However, the "recent migrant" and "frustration-aggression" theories discussed here are the two theories that appear most often in discussions of urban problems in the developing nations.

The threat to political stability lies not with newcomers, but with those deprived and frustrated slum and shanty dwellers who are longer urban residents, or with the second generation. These contentions are explored in turn.

#### THE DISRUPTIVE MIGRANTS

*The theory.* In cities growing at annual rates of 5 percent or more, with migration accounting for more than half the increase, migrants come to constitute the bulk of the urban population within a few decades. Moreover, many are very recently arrived: a quarter or more of the population of many major Latin American cities have arrived within the past five years.<sup>6</sup> It is often assumed that many or most of the newcomers have been torn from a tightly structured rural society and plunged into a bewildering, impersonal, and harsh environment with few or no sources of support and guidance. Shock and isolation produce personal disorientation and political anomie. Economic conditions are also an acute disappointment. Frustrated and disillusioned, the migrant is readily persuaded to political violence or extremism.<sup>7</sup>

*The evidence.* There is virtually no evidence that new migrants are either radical or violence-prone. Direct evidence is hard to find, because it is difficult to separate newcomers from the rest of the population. Myron Weiner examined the Calcutta voting pattern for state assemblymen in 1957 and 1962, and found that districts with heavy concentrations of migrants (not necessarily newcomers) correlated closely with high votes for the Congress Party.<sup>8</sup> Soares and Hamblin, analyzing factors affecting the Chilean presidential elections of 1952, found a negative relation between departments with high proportions of migrants and a high vote for Salvador Allende, the candidate of the radical left.<sup>9</sup> Studies of recent urban violence in the United States have found migrants under-represented among the rioters.<sup>10</sup> Charles Tilly, studying urban turbulence in nineteenth-century France, finds no consistent relation at all between the frequency of violent conflicts in de-

<sup>6</sup> Bertram Hutchinson, "The Migrant Population of Urban Brazil," *America Latina*, vi (April-June 1963), 43, 46; Paul T. Schultz, 2.

<sup>7</sup> Philip Hauser, "The Social, Economic, and Technological Problems of Rapid Urbanization," in Bert F. Hoselitz and Wilbert E. Moore, eds., *Industrialization and Society* (The Hague 1963), 210-11; Glaucio Soares and Robert L. Hamblin, "Socio-economic Variables and Voting for the Radical Left, Chile, 1952," *American Political Science Review*, lxi (December 1967), 1055. See also Mancur Olson, "Economic Growth as a Destabilizing Force," *Journal of Economic History*, xxiii (December 1963), 534.

<sup>8</sup> Myron Weiner, "Urbanization and Political Protest," *Civilisations*, xvii (1967).

<sup>9</sup> Soares and Hamblin.

<sup>10</sup> National Advisory Commission of Civil Disorders, *Report*, advance edition printed by The New York Times Company, 1968, 130-31; Lupsha, 7.

partments and the rates of increase of the urban populations in the cities dominating those departments.<sup>11</sup>

*Where the theory goes wrong.* Data that shed light on where the theory goes wrong are more plentiful than direct evidence on the political behavior of newcomers to the cities.

1. Regarding the shock of urban life: Many migrants into the great cities come from smaller cities and towns, not from the countryside. Surveys indicate that in Santiago, Chile,<sup>12</sup> and in six Brazilian cities including Rio and São Paulo,<sup>13</sup> roughly two-thirds of the migrants were *not* rural in origin. This reflects the well-documented pattern of step-wise migration. In India, higher proportions of migrants come directly from the countryside. Surveys in the mid-1950's found that 50 and 75 percent of the migrants to Delhi<sup>14</sup> and Bombay,<sup>15</sup> respectively, were rural in origin.

Among those migrants who do come from small villages, many have some previous exposure to urban influence. Surveys in squatter settlements in Bogotá<sup>16</sup> and Ankara<sup>17</sup> found that two-thirds of the migrants came from less than 100 and less than 120 miles away, respectively. Villages near large cities are often affected by urban influences. Many migrants have also visited city relatives or friends before deciding to move. There are villages where for many years high proportions of the young people have migrated; earlier migrants return on visits to tell prospective migrants what they can expect.

Particularly in Latin America, traditional rural social structure has been eroded in all except the most remote areas.<sup>18</sup> Therefore many of those migrants who are rural in origin do not come from tightly structured social settings. Though rural change and disorganization have little to do with affirmative preparation for urban life, these trends do cast doubt on the assumption that migrants are traumatized by the shattering of deeply ingrained habits and values.

<sup>11</sup> Charles Tilly, "Urbanization and Political Disturbances in Nineteenth-Century France," mimeo., presented to the annual meeting of the Society for French Historical Studies (Ann Arbor, April 1966), 7-8.

<sup>12</sup> Herrick, 53-103.

<sup>13</sup> Hutchinson, 43-44.

<sup>14</sup> V. K. R. V. Rao and P. B. Desai, *Greater Delhi: A Study in Urbanization 1940-1957* (Bombay 1963), 159.

<sup>15</sup> D. T. Lakdawala, *Work, Wages, and Well-being in an Indian Metropolis: Eco-1957* (New York 1965), 79.

<sup>16</sup> William L. Flinn, "Rural-to-Urban Migration: A Colombian Case," Research Publication No. 19 (U. of Wisconsin Land Tenure Center July 1966), 10, 23.

<sup>17</sup> Granville Sewell, *Squatter Settlements in Turkey*, unpub. diss. (M.I.T. 1964), 304.

<sup>18</sup> Marshal Wolfe, "Some Implications of Recent Changes in Urban and Rural Settlement Patterns in Latin America," paper presented at the U.N. World Population Conference (Belgrade, September 1965), 25.

Finally, since most migrants are young adults with better education and training than the average in their places of origin, their capacity and desire to adjust can be presumed to be comparatively high.

2. Regarding isolation: Newcomers obviously know few people when they arrive in the city. But very high proportions—from 70 to 90 percent—of samples drawn in several Latin American cities<sup>19</sup> report receiving help in settling from family, friends, or employers. In some regions and countries, new arrivals may also find assistance and friends in provincial, home-town, tribe, or caste associations.<sup>20</sup>

3. Regarding economic conditions and migrants' reaction to them: Most migrants find jobs quite quickly. (A Santiago sample: 40 percent within 2 days;<sup>21</sup> a second Santiago sample drawn from squatters only: 47 percent within the first week.<sup>22</sup> A Brazilian sample of the adult population in six cities: 80 percent within a month.<sup>23</sup> Other surveys report similar findings.)<sup>24</sup> Surveys also consistently show lower rates of open unemployment among migrants than among native urbanites.<sup>25</sup> This, however, may reflect age structure: higher proportions of native-born members of the labor force are adolescent or in their early twenties, age groups with higher-than-average unemployment rates.<sup>26</sup> Migrants are somewhat over-represented in low-level jobs, but the overall patterns of employment are not markedly different from those of urban-born workers.<sup>27</sup> This probably reflects the high proportions of migrants who come from other cities and towns, and who may well be better educated and more highly skilled than the average for the city of destination. The few surveys that compare current jobs (or

<sup>19</sup> Gino Germani, "Inquiry into the Social Effects of Urbanization in a Working Class Sector of Greater Buenos Aires," United Nations Economic and Social Council, E/CN.12/URB/10, December 1958, Table 10, p. 26; Herrick, 91, and ECLA, "Urbanization in Latin America: Results of a Field Survey of Living Conditions in an Urban Sector," mimeo., E/CN.12.622, 1963, 17; Flinn, 27; Hutchinson, Table 12, p. 61.

<sup>20</sup> William Mangin, "The Role of Regional Associations in the Adaption of Rural Migrants to Cities in Peru," in Dwight Heath and Richard Adams, eds., *Contemporary Cultures and Societies of Latin America* (New York, 1965), 319.

<sup>21</sup> Herrick, 92.

<sup>22</sup> ECLA, "Results of a Field Survey," 16.

<sup>23</sup> Hutchinson, 67-68.

<sup>24</sup> Germani, 69, Table 37.

<sup>25</sup> Slighton, 38; Herrick, 84; Weiner, "Urbanization and Political Protest," 6; Rao and Desai, 341, Table 16-1, and 383, Table 17-3; Dantwala, 481; R. Mukerjee and B. Singh, *Social Profiles of a Metropolis* (Bombay 1961), 116; G. M. Farooq, *The People of Karachi: Economic Characteristics*, Monographs in the Economics of Development No. 15 (Karachi, Pakistan Institute of Development Economics, July 1966), 19.

<sup>26</sup> Herrick, 79; Slighton, 37.

<sup>27</sup> Germani, 51, Table 29; Rao and Desai, 373, Tables 16-19; Herrick, 86-87, Tables 6-9; Lakdawala, 466, Table VI-37, Columns 7-8; Mukerjee and Singh, 88-89; Rao and Desai, 223, Tables 12-15.

first job in the city) with jobs before migration show considerable upward mobility.<sup>28</sup>

Migrants' reactions to city housing vary, depending on their previous standards and on the geographic and climatic conditions that largely determine the objective adequacy of low-cost housing and shanty settlements. Germani's working-class samples in Buenos Aires, most of whom came from smaller cities or towns, were quite dissatisfied.<sup>29</sup> But Pearce and Bonilla both stress that *favela* housing in urban Brazil is much like rural housing throughout the nation.<sup>30</sup>

More generally, surveys conducted in settlements ranging from well-established blue-collar neighborhoods in Buenos Aires to squalid shantytowns outside of Baghdad concur: overwhelming proportions of migrants say that their incomes, material possessions, access to services, and opportunities for their children are better in the city.<sup>31</sup>

To summarize: The assumptions that migrants are uprooted and isolated in the city are grossly overdrawn. The assumption that most are disappointed and frustrated by economic conditions is simply wrong. Some migrants undoubtedly are disillusioned, but lack of widespread contacts plus political inexperience and traditional patterns of deference makes it most unlikely that newcomers' frustrations will be translated into destabilizing political action.

#### THE RADICAL MARGINALS

*The theory.* To exorcise the myth of the disruptive migrants is to raise a new specter—that of the radical poor. If migrants feel an initial sense of progress, what happens when their memories of earlier misery fade? If low levels of political awareness, deference to authority, and perhaps political conservatism are part of the rural baggage migrants carry with them, what happens after prolonged urban exposure? As migrants become established urbanites with urban aspirations and attitudes, but still eke out a marginal existence, will not they or their children sooner or later express growing frustration through political radicalism or violence?<sup>32</sup>

<sup>28</sup> Rao and Desai, 94-95, Tables 5-14, 5-15; Herrick, 94-95.

<sup>29</sup> Germani, 16.

<sup>30</sup> Andrew Pearce, "Some Characteristics of Urbanization in the City of Rio de Janeiro," in Philip Hauser, ed., *Urbanization in Latin America* (UNESCO 1961), 196; Frank Bonilla, "Rio's Favelas: The Rural Slum Within the City," *American University Field Staff Reports*, East Coast South America Series, 8:3 (1961), 2.

<sup>31</sup> Doris Phillips, "Rural-to-Urban Migration in Iraq," *Economic Development and Cultural Change*, vii (July 1959), 417; Flinn, 5, 37; Sewell, 109-110.

<sup>32</sup> For example, see Glaucio Soares, "The Political Sociology of Uneven Development in Brazil," in Irving L. Horowitz, ed., *Revolution in Brazil* (New York 1964), 192,

*The evidence.* The theory is hard to test empirically. Data on length of urban residence are scarce. Adequate indices of radicalism and workable definitions of marginality are difficult to devise. Glaucio Soares states the theory particularly clearly but cites only one piece of evidence: a survey conducted among skilled and unskilled workers in Rio in 1960, showing that support for the Brazilian Labor Party increased substantially among unskilled workers with longer urban residence, but not among skilled workers.<sup>33</sup> I have tried to approximate his test using data from the Almond-Verba surveys in Mexico<sup>34</sup> and from Inkeles' surveys in Argentina and Chile.<sup>35</sup> The Mexico data contradicts Soares' Rio findings: skilled workers with greater urban experience leaned toward demonstrations and violence as effective political techniques; unskilled workers did not. The Inkeles data looks more like Soares' findings, but the index of radicalism is inadequate: replies to questions regarding the need for substantial and rapid social change, with no reference to political means.

Studies of class voting patterns and of participation in urban violence do not distinguish among degrees of urban experience and therefore do not really test the hypothesis. But in the absence of good tests, such studies at least provide data on the political behavior of the urban poor in general. In the Venezuelan presidential elections of 1958 and 1963, there was little support in the poorest districts of Caracas for the Communist ticket in 1958 and the farthest left legal candidate in 1963. In Santiago, Chile, high proportions of the low-rent district vote went to Salvador Allende in 1958, but in 1964 much of this support shifted to Eduardo Frei, a moderate reformist alternative not previously available.<sup>36</sup> Indian survey data gathered before the 1966 national elections found low-income voters in Bombay, Calcutta, Delhi, and Madras no more and no less ready to support extremist parties of the right and left than were more affluent voters.<sup>37</sup> Myron Weiner states that demon-

---

195; also Kingsley Davis and Hilda H. Golden, "Urbanization and the Development of Pre-industrial Areas," *Economic Development and Cultural Change*, III (1954), 19-20.

<sup>33</sup> Soares, 192.

<sup>34</sup> Gabriel Almond and Sidney Verba, *The Civic Culture* (Princeton 1963).

<sup>35</sup> Large-scale attitudinal surveys of factory workers and control groups were conducted by Alex Inkeles and his associates in Argentina, Chile, Israel, Nigeria, India, and Pakistan, as a basis for a study of the modernizing impact of factory experience on attitudes. I am indebted to Professor Inkeles for permission to use his data.

<sup>36</sup> George F. Jones, "Urbanization and Voting Behavior in Venezuela and Chile, 1958-1964," typescript prepared at Stanford University, March 1967, 40-43, 69-72.

<sup>37</sup> Indian Institute of Public Opinion, "The Structure of Urban Public Opinion," *Public Opinion Surveys*, XI (February 1966), 15-16.

strations in Calcutta are much more likely to involve violence if they are based on the middle class than if they are predominantly working class.<sup>38</sup> Tilly and other students of French history conclude, from formidably detailed police records, that those involved in the repeated Parisian incidents of the nineteenth century were mostly from the skilled, established crafts, "segments of the working class already politically alert, organized, and integrated into the life of the city."<sup>39</sup>

*Where the theory goes wrong.* The radicalization theory has more validity than the myth of the disruptive migrants, but is seriously overstated.

1. Regarding the aspiration-achievement gap: Aspirations undoubtedly rise, but they probably do so quite gradually. No studies are available, to my knowledge, on the effect of urban exposure on material and other aspirations. My own attempts to use Inkeles' data for this purpose do show an upward drift in aspirations as urban experience lengthens. Rising aspirations for better jobs, housing, and status are more marked, however, among the skilled than among the unskilled workers. The urban-born are less satisfied with their jobs and self-assessed status.<sup>40</sup>

Sociological studies of working-class norms provide a firmer basis for the hypothesis that aspirations rise slowly. Those at the very bottom are preoccupied with survival. Some may be victims of the apathy Oscar Lewis calls "the culture of poverty." Those a step or two up the ladder set modest targets for themselves: sons of unskilled workers in Querétaro, Mexico, want to be masons, shoemakers, or construction workers.<sup>41</sup> Realism may not only trim targets but may also shape values. Studies of working-class norms in the United States find "getting by" stressed more heavily than "getting ahead."<sup>42</sup>

Turning to achievement: modest occupational progress out of the lowest categories is probably quite widespread. Data on intergenerational occupational mobility in three cities in developing countries—São Paulo, Buenos Aires, and Poona—show a majority of the sons of unskilled workers rising out of this category, and substantial fractions

<sup>38</sup> Myron Weiner, "Violence and Politics in Calcutta," *The Journal of Asian Studies*, xx (May 1961), 277.

<sup>39</sup> Charles Tilly, "À Travers le Chaos des Vivantes Cités," mimeo. paper presented to the Sixth World Congress of Sociology (Evian-les-Bains, September 1966), 17, 19.

<sup>40</sup> Joan Nelson, 45-51.

<sup>41</sup> Andrew Whiteford, *Two Cities of Latin America: A Comparative Description of Social Classes* (New York 1964), 120.

<sup>42</sup> S. Michael Miller and Frank Reisman, "The Working-Class Sub-culture," *Social Problems*, ix (Summer 1961), 92, 95-96; Herbert J. Gans, *The Urban Villagers* (New York 1962), 219-221, and passim.



rising into nonmanual occupations<sup>43</sup> (Table 1). São Paulo may be uniquely dynamic, but the data from Poona probably is not unrepresentative. Moreover, Germani's data from Buenos Aires show that the second generation, with its higher aspirations, is also more likely to move out of the unskilled category.<sup>44</sup>

The much-deplored squatter settlements may provide many urban poor with an additional channel of achievement. Such settlements are

TABLE I  
OCCUPATIONAL MOBILITY OF MEN WHOSE FATHERS WERE UNSKILLED

| Occupational level of respondents   | São Paulo, Brazil (1960)* | Buenos Aires, Argentina (1960)* | Poona, India (1954)* |
|---|---------------------------|---------------------------------|----------------------|
|   | %                         | %                               | %                    |
| 1. Unskilled manual workers, domestics, vendors.  | 31.2                      | 21.1                            | 43.2                 |
| 2. Semiskilled and skilled manual workers, including supervisory manual grades.                             | 42.0                      | 52.6                            | 26.1                 |
| 3. Lower-level nonmanual, including supervisors, shop and office clerks, small businessmen (1-5 employees). | 12.1                      | 21.1                            | 22.2                 |
| 4. Higher-level nonmanual and owners of medium-small businesses.  | 8.6                       | 2.3                             | 3.2                  |
| 5. Administrative and executive posts; owners of medium-large stores and factories; liberal professions.    | 3.0                       | 2.9                             | 1.0                  |
| 6. High professional and administrative posts.  | 3.0                       | 0.0                             | 0.8                  |

\* The samples include: São Paulo: 338 sons of unskilled fathers from a total sample of 1056 adult men; Buenos Aires: 341 sons of unskilled fathers, from a total sample of 2078 heads of households; Poona: 2004 sons of unskilled fathers, from a total sample of 4505 heads of households.

<sup>43</sup> Sources for data in Table I: São Paulo: Bertram Hutchinson's survey data as reported in S. M. Miller, "Comparative Social Mobility," *Current Sociology*, ix (1960), 69. Buenos Aires: Gino Germani, "La Movilidad Social en la Argentina," mimeo., Publicación Interna No. 60, Instituto de Sociología, Facultad de Filosofía y Letras, Universidad de Buenos Aires. Poona: N. V. Sovani, "Occupational Mobility in Poona Between Three Generations," in *Urbanization and Urban India* (Bombay 1966), 96.

<sup>44</sup> Gino Germani, data developed for but not presented in the study on social mobility cited in footnote 43. I am indebted to Professor Germani for permission to use his data.

not usually the first stop for the incoming migrant, but are more often established by families with longer urban experience. Where squatters are not harassed by the police, and where terrain and initial density of settlement permit, many shantytowns evolve over ten or fifteen years into acceptable working-class neighborhoods.<sup>45</sup> Squatting offers immediate relief from the burden of rent and the threat of eviction, and a long-run prospect of a modicum of comfort and respectability. In many cities in Latin America, North Africa, and Turkey, 10 to 40 percent of the population live in squatter settlements.<sup>46</sup> Not all of these settlements are "self-improving," but for many among the urban poor, squatting is a means of substantial progress.

2. Regarding frustration: The slow rise of aspirations and the wide incidence of modest progress may explain the otherwise puzzling finding that even among groups reporting that their economic situation has deteriorated in the past few years, overwhelming majorities express optimism about the future<sup>47</sup> and belief that the economic and social system is open to talent and hard work. Surveys in many different countries, asking a variety of questions, concur on this point.<sup>48</sup> Faith in the future may make present hardship less bitter. And general belief in open opportunity prevents conversion of individual misfortune into a sense of social injustice.

3. Regarding conversion of frustration to disruptive political action: Frustration obviously can be expressed in many ways—withdrawal and defeat, quarrelsome behavior, alcoholism, religion. Political ac-

<sup>45</sup> See John Turner, *Uncontrolled Urban Settlement: Problems and Policies*, Working Paper No. 11 for the Inter-Regional Seminar on Development Policies and Planning in Relation to Urbanization, organized by the United Nations Bureau of Technical Assistance Operations and the Bureau of Social Affairs, October-November 1966, paper numbered 67-44032; Daniel Goldrich and others, "The Political Integration of Lower Class Urban Settlements," mimeo., prepared for the American Political Science Association meetings, September 1966, 4; Flinn, 3-4; William Mangin, "Latin American Squatter Settlements: A Problem and a Solution," *Latin American Research Review*, 11 (Summer 1967), 74-75.

<sup>46</sup> Estimates place a fifth to a quarter of Lima's population in the early 1960's in squatter settlements; 16% of Rio as of 1964; 30% of Caracas in the late 1950's (despite construction of immense public housing projects absorbing an additional 18% of the city's population); and over a third of Mexico City. Richard M. Morse, "Recent Research on Latin American Urbanization," *Latin American Research Review*, 1, 1965, 50; John Turner, 1. In Turkey, Granville Sewell estimates that squatters comprise a fifth of Istanbul, a third of Ankara, and a third of Adana. 71, 186, 193.

<sup>47</sup> Santiago and Lima: Goldrich. Caracas: CENDES print-out. Mexican working class: data drawn from taped data of interviews conducted by Gabriel Almond and Sidney Verba for their study of the "Civic Culture" in Mexico and other countries.

<sup>48</sup> Bonilla, 11; CENDES print-out; Goldrich, "Politics and the Pobladores," Tables 1, 17, and "Demographic and Socio-economic Background, Social Mobility, and Expectations"; Germani, "Social and Political Consequences," 389-90. Inkeles' survey data demonstrate a similar belief in the hard-working poor man's prospects. See Nelson, 61.

tion, whether individual or associational, moderate or extremist, legal or illegal, is only one class of reactions among many. The point is obvious, but it is often overlooked.

Moreover, the urban poor are particularly unlikely to choose aggressive political action to express their grievances. Low status and lack of contacts make them more vulnerable to police force and less likely to get fair treatment in the courts. They are unorganized, their level of political awareness is low, and these conditions are perpetuated by the constant inflow of additional unskilled, poorly educated migrants. Moreover, the causes of poverty, unemployment, and inflation are remote. Wealthier and better-educated people may think in terms of governmental policies and their effect on economic conditions, but the poor and uneducated are less likely to blame the authorities for general economic difficulties<sup>49</sup> (Table 2). The recent militancy of minority groups among the poor in the United States appears to contradict these generalizations, but that is the product of very different circumstances.

To summarize: The theory of radicalization and the deprivation/frustration/aggression model on which it rests may contain a good deal of truth. The theory itself does not specify the pace of radicalization. But it is usually linked to predictions of political upheaval in the not-too-distant future. This implicit estimate almost surely overstates the rate at which both discontent and political awareness spread among the urban lower classes. Such calculations also imply a high rate of conversion from frustration to radical political action. In fact, the bulk of economic and social frustration is likely to leak into alternative channels, including nonradical political action.

The question of destabilization or radicalization has pre-empted much of the meager attention scholars have given to the study of the political implications of urban growth abroad. Yet if we want to locate the sources of urban turbulence and extremism, the urban lower class is the wrong group to examine. Indeed, most of those scholars primarily interested in civil conflict have focused their attention on other groups, primarily students, organized labor, and the military. If, however, we are interested in the current and potential political role of the urban lower class, preoccupation with instability blinds us to a number of more probable, and equally interesting, patterns of evolution.

In brief, most inquiries into the political role of the urban poor in developing countries have asked the wrong questions. Moreover, most analysis to date has stressed social psychology to the point of ignoring

<sup>49</sup> Derived from Indian Institute of Public Opinion, *Public Opinion Surveys*, x (March 1965), 21-32.

TABLE 2

"To whose policy do you ascribe this shortage (of foodgrains)?"

|  | No. reporting<br>shortfall | % blaming<br>government | % blaming<br>others | % blaming<br>both |
|--|----------------------------|-------------------------|---------------------|-------------------|
| <i>Income Group</i>                                    |                            |                         |                     |                   |
| Under Rs. 75   | 30                         | 30                      | 23                  | 30                |
| 76-150   | 142                        | 58                      | 12                  | 23                |
| 151-300  | 182                        | 59                      | 9                   | 31                |
| 301-500  | 67                         | 58                      | 6                   | 30                |
| <i>Education</i>                                       |                            |                         |                     |                   |
| Illiterate and<br>literate but with-<br>out formal ed. | 56                         | 37                      | 7                   | 30                |
| Primary, under-<br>matric. and<br>matric.              | 229                        | 49                      | 8                   | 35                |
| Some college   | 270                        | 59                      | 8                   | 28                |
| <i>Occupation</i>                                      |                            |                         |                     |                   |
| Worker   | 38                         | 42                      | 16                  | 34                |
| Salaried em-<br>ployee                                 | 265                        | 57                      | 9                   | 26                |
| Independent  | 163                        | 60                      | 8                   | 30                |

Source: Indian Institute of Public Opinion, all-urban sample of 1000, March, 1965. Differences between totals and 100% are "don't know."

highly relevant political variables. The attitudes of the urban poor, so far as they can be ascertained, are certainly important. But these attitudes are shaped by and interact with the political context in the individual nation and city. The agents and channels of political socialization of the urban poor, and the factors determining the degree and nature of interest in the urban poor on the part of established political organizations, are topics that deserve much more attention than they have received thus far.

## II. THE CAPACITY OF THE URBAN POOR FOR POLITICAL PARTICIPATION

Built into the lives of the urban poor are a multitude of factors impeding political mobilization. Industrial labor and workers in large modern service establishments are relatively well-paid and secure, usually unionized, and often politically active. In contrast to this labor elite, the bulk of the urban poor work in occupations that provide few handles for political mobilization. Small artisan and service shops,

vending, and domestic service are inherently difficult to organize. Stevedoring and construction are organizable, but the relentless pressure of unskilled job-seekers impedes unionization. Even social and religious organizations are weak among the urban poor. Insecure jobs and rented rooms mean that many move frequently, which probably prevents the growth of interest in and knowledge of local issues, certainly impedes the growth of ties within the neighborhood, and in the past has legally disenfranchised large numbers of the poor. In many cases, ethnic and religious cleavages further reduce the potential for collective action.

Moreover, the most urgent needs of the very poor are intensely individual—a day's work, a room, a week's credit, medical aid, someone to fix trouble with the police. Others equally poor and ill-connected usually cannot offer much help. The poor, rural as well as urban, turn for aid to better-established individuals or organizations—landlord, employer, shopkeeper, church, settlement house, regional or caste or tribal association, local fixer, ward boss. The patron-client relations that result may or may not form a stable pattern. They are less likely to do so in an urban setting where there are several possible sources of assistance, and where the best or most accessible source of help for one problem may be less appropriate or available for other problems. Where a stable pattern exists, it may or may not be tied into the broader political system, depending on the interests and connections of the patron(s). Politicized patron-client relations therefore are possible, but not automatic, in an urban setting.

In the cities of the United States during the late nineteenth and early twentieth centuries, both the needs of the urban poor and their ethnic loyalties were used as bases for powerful political machines. But while their votes might elect the candidates of the machine, the urban poor had little or no influence over the city governments they put in office. The nature of their political role—the large-scale institutionalized trading of individual votes for individual favors—precluded any substantial impact on governmental programs and policies. Political integration in this more fundamental sense of influence on governmental actions took place, in the experience of Europe and North America, only as the urban poor were gradually absorbed economically and socially, that is, as they joined the stable and organized working class.

In the cities of today's developing nations, the social and economic characteristics of the urban poor are in many ways similar. Is there then any reason to expect them to play a different or more active or influential political role in Latin America, Africa, and Asia than they

played historically in the industrialized nations? Perhaps not. But there are major differences as well as similarities between current and past patterns of urban growth and industrial expansion. These contrasts may produce different political results.

First, cities today are growing much more rapidly. Annual growth rates of 5 to 8 percent are common among the cities of the less developed world. In contrast, few major European cities grew more rapidly than 3 or 4 percent a year during any decade of the nineteenth century.<sup>50</sup> North American cities mushroomed rapidly in their early stages but then slowed down, despite heavy immigration as well as internal migration. Today's more rapid growth in the developing world results not from heavier migration into the cities but from lower urban mortality rates, in turn the product of improved public health measures.<sup>51</sup> Moreover, in Europe millions chose to emigrate rather than move to their own growing cities, and in North America the frontier long offered an alternative channel for the restless and ambitious. In the developing countries today, few can emigrate, and only a handful of the nations of Latin America and Africa have extensive arable frontier regions. In most countries there is only one magnet for those who want a better life: the city.

Second, although cities are growing faster, urban jobs are probably multiplying less rapidly than they did in the past. Surely this is true of employment in manufacturing. Even in those few centuries where industry is making good progress, employment in manufacturing rarely expands by more than 3 or 4 percent a year.<sup>52</sup> Modern technology is, of course, laborsaving. Moreover, poorly trained and unreliable labor plus legally required social security benefits may mean that labor costs per unit of output are high despite low hourly or daily wages. Development policies such as rapid write-off and other tax concessions and special exchange rates for imported capital equipment reduce the cost of capital compared to labor. From the entrepreneur's standpoint, labor may be relatively costly and capital comparatively cheap even in countries where labor is plentiful and capital scarce.

In short, demographic and economic trends conspire to produce a larger marginal class persisting over a longer period than in the cities

<sup>50</sup> Adna F. Weber, *The Growth of Cities* (New York 1899), chap. 2.

<sup>51</sup> Kingsley Davis, "The Urbanization of the Human Population," in *Cities* (New York 1966), 18-19.

<sup>52</sup> Economic Commission for Latin America (ECLA), "Structural Changes in Employment within the Context of Latin America's Economic Development," *Economic Bulletin for Latin America*, x (October 1965), 166; see also Werner Baer and Michel Herve, "Employment and Industrialization in Developing Countries," *Quarterly Journal of Economics*, LXXX (February 1966).

of nineteenth- and early twentieth-century Europe and North America. In view of the limited capacity of the urban poor for political action, a large mass of marginals persisting over many decades has no necessary political implications. However, the political climate and structure of today's developing nations differ from those of nineteenth-century Europe and North America in ways that may produce more political activity among the urban poor.

The poor themselves may be better educated and informed, and therefore may have a greater capacity for political action than did their historical counterparts. Because urban mortality rates are much lower, a higher proportion are urban-born. In the cities, free or cheap primary schools are more widely available than they were in the late nineteenth or early twentieth century in Europe and North America. Radio now reaches even the illiterate.

For many of the urban poor, squatting also provides a social setting that is more conducive to collective action. Squatting was not unknown in Europe and North America, but is far more extensive in today's developing countries, largely as a result of more rapid urban growth. Widespread squatting almost certainly implies more stable residence patterns. Stephan Thernstrom<sup>53</sup> and other historians interested in United States social history around the turn of the century report incredibly high rates of turnover among the urban poor. Although housing is only part of the picture, it is clear that a man who rents his shelter and loses his job is likely to be evicted, while a man who owns his house and loses his job is likely to cling to his remaining source of security.

Stable residence and home ownership or quasi-ownership encourage collective action for collective neighborhood interests. Studies of squatter settlements often mention neighborhood councils that are more or less active and effective in pressing for legal recognition and extension of city services into the community.<sup>54</sup> Their demands are concrete and often limited: for example, the covering of a former irrigation ditch that endangers neighborhood children. Some goals are startling for their upward-striving character: Alejandro Portes describes a government-sponsored community in Santiago, Chile, where the *pobladores* insisted that their monthly payments for their lots and houses be in-

<sup>53</sup> Stephan Thernstrom, "The Case of Boston," *Massachusetts Historical Society*, (Autumn 1967), 114-15, and "Working-Class Social Mobility in Industrial America," mimeo., prepared for delivery at the Anglo-American Colloquium of the Society for Labour History (London, June 23, 1968), 5-6.

<sup>54</sup> William Mangin, "Latin American Squatter Settlements," 69-70; Talton Ray, *The Politics of the Barrios of Venezuela* (Berkeley 1968), *passim*.

creased to hasten the day when they could claim clear ownership.<sup>55</sup> Nonetheless the combined demands of many neighborhoods can seriously strain the administrative and financial capacity of responsive city or national governments. From an economic standpoint, therefore, it is important to add that many communities have proved capable of substantial self-help programs, contributing funds and labor for a series of community improvements with or without external support. In some cities, presumably where squatters' relationship to legal authorities is unclear or insecure, neighborhood councils have also played a quasi-judicial role, adjudicating local disputes regarding boundaries, right of way, and similar issues.<sup>56</sup> Neighborhood councils undoubtedly are providing many of the urban poor with their first experience in limited local self-government.<sup>57</sup> Recognizing this potential, Chile in 1968 passed legislation that grants legal status to the neighborhood councils that had grown up spontaneously in many local-income urban areas, gives them a voice in the allocation of the (very limited) annual municipal budgets, authorizes them to borrow from the banks, and empowers them to share in other ways in urban government.

Neighborhood councils in low-income urban areas also offer political parties an organizational channel through which to reach the urban poor on a more sustained, institutionalized basis than periodic campaign tours and speeches. Whether parties take advantage of the opportunity depends largely on the national political context—on the importance of elections, the extent of the franchise, the degree of party competition, the nature of major parties' ideologies and leadership, and other factors, all of which vary tremendously. However, the political context of urban growth in Latin America, Africa, and Asia appears more likely to encourage responsiveness to or alliances with the urban poor than was the case in the now-industrialized nations in the late nineteenth and early twentieth centuries. The franchise is considerably broader, if only because women are included in the electorate in almost all developing countries. Despite the immense diversity of political systems, ideas about the proper role of government and the scope of governmental responsibility for the welfare of its citizens are every-

<sup>55</sup> Alejandro Portes, "Los Grupos Urbanos Marginados: Un Nuevo Intento de Explicación," typescript (Santiago, Chile, June 1969).

<sup>56</sup> Kenneth Karst, "Preliminary Report on a Study of the Internal Norms and Sanctions in Ten Barrios of Caracas," mimeo., talk prepared for a meeting of Latin American Scholars (New York, Autumn 1968).

<sup>57</sup> This experience is not necessarily positive. See Daniel Goldrich and others, "Political Integration," 10-14.



where more interventionist than in even the most progressive nations fifty years ago.

In sum, while the urban poor are unlikely to play a major political role in the developing countries, they may well prove more active than were their counterparts in the past experience of the industrialized nations. Urban marginals are likely to be more numerous relative to total urban population. They may also prove somewhat more alert and sophisticated, with greater capacity for political organization. This capacity will probably be matched, in at least some countries, with stronger interest on the part of established political groups in organizing and tapping the political potential of the poor.

### III. POSSIBLE PATTERNS OF POLITICAL INTEGRATION

The suggestions that follow are doubly speculative. Any effort to anticipate the pattern of events as they unfold is necessarily speculative. But even early returns can provide useful clues to eventual outcomes. Hence my second disclaimer: what follows are preliminary hypotheses, supported by general structural considerations but not yet tested against available evidence.

I start from the assumption that in most developing countries the urban poor are at best weakly integrated politically. This need not mean that they are alienated from the system. It may mean merely that few among them have much interest in politics, or perceive politics as relevant to their interests. They may regard the government as legitimate (or at least not illegitimate) but (probably quite realistically) view politics and politicians cynically, and discount the possibility that they themselves can influence government policies or actions. Political parties, for their part, may give lip service to the needs of the urban poor, and candidates will put in an appearance in slum and squatter areas during political campaigns. One or more parties may make a more serious effort to organize support based on patron-client relations. However, they are unlikely to alter their programs significantly, nor, if they take office, to reflect the needs of the poor in budget priorities. The exchange of individual favors for individual votes falls far short of any sustained influence on party priorities or programs.

Sharp ethnic cleavages can change this picture drastically. The easiest and in some cases the only effective way for elite or middle-class leaders in racially divided societies to expand their following is to appeal across class lines to ethnic loyalties. Race or religion transforms legitimacy into a simple, concrete, and highly charged issue, and also provides a ready link between legitimacy and bread-and-butter issues such

as jobs. Race also provides a basis on which to appeal simultaneously to urban and rural groups whose interests and outlooks might otherwise conflict. In short, race politics operates as a powerful unifying and mobilizing force within racially homogeneous segments of the urban lower class and across class and rural-urban boundaries. However, it may be profoundly disintegrative at the national level, creating or exacerbating tensions as disruptive as and more enduring than those based on class in the absence of race cleavage.

In racially homogeneous societies, class is the more formidable barrier to political integration. In theory, one possible pattern of evolution in such a society might be the formation of class consciousness and class-based political organization. Some analysts of Latin American politics, particularly those with European backgrounds, seem to regard this possibility as not unlikely. Several surveys have attempted to detect the emergence of class consciousness among factory workers or among more marginal groups.<sup>58</sup>

Such a pattern seems to me improbable. The highly individual needs of the very poor plus their distrust and lack of organizational experience militate against collective organization. Kerr and Siegel's theory of isolated masses suggests that shared work experience and conditions of life, coupled with isolation from the broader society, are likely to produce unified and aggressive action.<sup>59</sup> But by these criteria the urban poor in today's developing nations are less likely to develop strong group consciousness than were those of yesterday's European and North American cities. A smaller proportion share the common experience of factory work; more earn a precarious living in constantly shifting day labor, with little opportunity to develop ties with their fellow workers or hostility toward a common boss. Fewer live in rented tenement rooms in solidly lower-class neighborhoods; more are housed in squatter settlements where they may have as neighbors (though not necessarily friends) aspiring lower-middle-class clerks, policemen, or teachers. Maurice Zeitlin's study of the determinants of revolutionary attitudes among Cuban workers suggests that those in occupations in which management and workers are sharply separated and have little contact are most likely to develop radical sentiments.<sup>60</sup> But many mar-

<sup>58</sup> See, for example, Alain Touraine, "Conscience ouvrière et développement économique en Amérique Latine," *Sociologie du Travail*, ix (July 1967), 229-54.

<sup>59</sup> Clark Kerr and Abraham Siegel, "The Inter-Industry Propensity to Strike: An International Comparison," in Arthur Kornhauser, Robert Dubin, and Arthur Ross, eds., *Industrial Conflict* (New York 1954).

<sup>60</sup> Maurice Zeitlin, *Revolutionary Politics and The Cuban Working Class* (Princeton 1967), chap. 6.

ginal occupations such as vending or day labor involve either no management or a constantly changing series of bosses, while domestic service and artisan manufacture are characterized by close and constant though not necessarily cordial relations with employers. Germani suggests that class consciousness in Latin America is also impeded by the constant inflow of rural migrants, diluting the shared experience and attitudes of the more experienced urban poor.<sup>61</sup>

Several alternative patterns seem much more probable than mobilization and organization of the urban poor on a class basis. One possibility is the emergence of a strong urban populist party appealing both to urban marginals and to industrial labor, possibly in alliance with low-level white-collar groups, on a platform stressing employment, public works, housing programs, and other immediate benefits. Such a pattern is more probable where urbanization and industrialization are comparatively extensive but established parties have proved unresponsive not only to urban marginals but also to industrial labor.

A second possibility is a pattern of gradually increased responsiveness to the needs (and votes) of the urban poor on the part of one or more of the established political parties, to the point of a significant revision of party programs and organization. Such a trend would appear most likely where elections are important in the national political system, party competition is substantial, and at least one major party is disposed to seek mass support. Interest in the urban poor may reflect ideology, a calculation of electoral advantage, or a combination of the two. In Italy in the early 1960's, the Communist Party recognized the political potential of mass migration from the south into the industrial cities of the north, reversed its earlier position endorsing restrictions on migration, and made an all-out play for the migrants' votes.<sup>62</sup> The Christian Democrats followed suit, but only after a lag. In Chile, also, the Communists and Socialists were first to try to capture local neighborhood councils in low-income areas of the cities; the Christian Democrats have since made a strong competitive bid. Where middle-class groups face an unyielding traditional elite, they may well seek alliance with the urban poor. In Venezuela, middle-class parties shut out of the political arena after a taste of power in the late 1940's found that they could use the power of the *barrios* to oust Pérez Jiménez in 1958. There-

<sup>61</sup> Gino Germani, "Social and Political Consequences of Mobility," in Smelser and Lipset, *Social Structure and Mobility in Economic Development* (Chicago 1966), 387.

<sup>62</sup> Robert C. Fried, "Urbanization and Italian Politics," *Journal of Politics*, xxxix (August 1967), 525.

after, they competed vigorously for control of local *barrio* organizations and first claim on the loyalties of the *barrio* residents.<sup>63</sup>

If it is the case that the urban poor in many developing countries have more capacity for semi-autonomous local organization for limited purposes than did their historical counterparts, the possibility of direct-confrontation tactics by organized groups of the urban poor is also greater than it was in the past. A particularly clear-cut and large-scale instance occurred in July, 1968, in Lima, where squatters threatened a march on the Presidential Palace unless the government decreed that they could be issued title to their lots.<sup>64</sup> With the proliferation of local neighborhood councils, mothers' organizations, and similar groups, it is easy to envisage more limited demonstrations along the lines of recent protests by welfare recipients in several cities in the United States. Direct pressure for needed services or immediately relevant policy changes is not necessarily an alternative to the other suggested patterns. It could well occur in combination with them. In dominant-party, single-party, or no-party states, however, petitions and limited local confrontations over specific issues may be the sole channels available, if indeed the authorities will tolerate this degree of pressure from a relatively weak group.

#### IV. CONCLUSION

Fear of the urban mob is as ancient as cities. Concern about rapid urbanization was widespread in Europe before and during the nineteenth century. But there is little ground for belief that the swelling urban masses of today's developing nations will prove to be politically radical or violent. There is, however, some reason to expect that they may play a more active and influential role than did their historical counterparts.

The important question, then, is how and under what conditions the urban poor are integrated into these political systems. In plural societies, race, religion, or tribe is the pole around which political organization will almost certainly coalesce. In more homogeneous societies, urban political machines may well appear, transforming traditional patron-client relationships into the large-scale institutionalized trading of votes for favors. Neighborhood associations in squatter settlements offer an alternative channel of political integration, with greater potential for influence by the poor on governmental activities.

<sup>63</sup> Ray, chap. 7.

<sup>64</sup> *New York Times*, September 24, 1968.

Or populist leaders may seek a coalition between poor urban voters and better-established working-class groups.

Any pattern of integration that gives the urban poor some significant influence on the allocation of material and human resources will alter the path (and may retard the pace) of economic growth. The political mobilization of the urban poor does, then, challenge the political systems of the developing nations. The challenge, however, is not the containment of extremist or anarchic outbursts, but the evolution of means to respond to concrete and usually moderate demands without the sacrifice of other development objectives.

## 21 The Absorption of Labor in the Urban Economy: The Case of Developing Countries

John Friedmann and Flora Sullivan

Cities in the developing countries are facing a problem of threatening dimension. Despite 2 decades of accelerated industrialization, often promoted through the familiar policy of import substitution, the rapidly increasing labor force of cities is not being absorbed into full, productive employment. Available statistics are unreliable and their coverage is incomplete. Such fragments as exist, however, suggest that low-productivity employment in nonagricultural occupations has been steadily rising on top of a layer of open unemployment that varies between 5 and 15 percent of the urban labor force.<sup>1</sup> The growing awareness of specialists that the employment situation in the developing countries was deteriorating led the International Labour Organization to initiate its World Employment Program in 1969. The aims of this program are to put an employment objective high on the agenda of national planners and to help them formulate and carry out employment-oriented development strategies.<sup>2</sup>

Two facts seem to account for the rising concern with labor absorption. The first is the continuing and even accelerating movement of people to cities. In the developing countries, urban growth rates are typically

<sup>1</sup> For an assessment of employment growth through 1980, see David Turnham and Ingelies Jaeger, *The Employment Problem in Less Developed Countries* (Paris: Development Centre, Organization for Economic Cooperation and Development, 1971), chap. v.

<sup>2</sup> International Labour Office (ILO), *The World Employment Programme*, Reports ILC/I/1 and ILC 56/IV (Geneva: ILO, 1969). The first report, by the director general of the ILO, lays the foundations for national and international action to make the creation of productive employment one of the main objects of general economic and social development policy. The second is a progress report on the world employment situation at the beginning of the United Nations Second Development Decade and describes the World Employment Program launched by the ILO and the kinds of activity carried out under it. Since 1969, the International Labour Office has published three major country studies on the employment situation: *Towards Full Employment: A Programme for Colombia* (Geneva: ILO, 1970); *Matching Employment Opportunities and Expectations: A Programme of Action for Ceylon*, 2 vols. (Geneva: ILO, 1971); and *Employment, Incomes and Equality. A Strategy for Increasing Productive Employment in Kenya*, 2 vols. (Geneva: ILO, 1972).

twice the rate of natural increase, and major cities have been known to double and even triple their populations within a single decade.<sup>3</sup> Assuming the rate of natural increase in cities to be equal to that of the nation as a whole, remaining urban growth can be attributed directly to migration.<sup>4</sup> And most of those who come to the city are young adults looking for work.

A second notable fact is that manufacturing industry, despite impressive gains in the value of its product, has not been opening up new jobs at anything like the rates required by the growth in urban labor. Between 1955 and 1965, and taking all developing countries together, industrial employment increased at an average annual rate of 4.4 percent (in Latin America the annual gain was only 2.5 percent), with per worker productivity rising, respectively, at 2.7 and 3.8 percent.<sup>5</sup> But in many cases, these percentage gains in employment were not only less than the corresponding rates of growth in the urban work force, but they also fell far short of the absolute number of jobs required.<sup>6</sup> Although the expansion of manufacturing was accompanied by the creation of additional employment in the "modern" sector, the great bulk of the manpower available to the urban economy was absorbed into small-scale enterprise, personal services, a nondefined "miscellaneous" category, and open unemployment.<sup>7</sup>

Because of these trends, the city in developing countries is perpetually struggling with crisis, even when both the direction and magnitude of change in national indicators, chiefly gross national product, are satisfactory. Despite substantial investments in urban housing, squatterments accumulate, and no amount of slum removal seems able to eradicate, once

<sup>3</sup> Relevant data are included in Kingsley Davis, *World Urbanization, 1950-1970*, vol. 1, *Basic Data for Cities, Countries, and Regions*, Population Monograph Series, no. 4 (Berkeley: Institute for International Studies, University of California, 1969), table D; and United Nations, *Growth of the World's Urban and Rural Population, 1920-2000* (New York: Department of Economic and Social Affairs, 1969).

<sup>4</sup> The range is typically from 33 to 75 percent of total population increase. For details, see World Bank, *Urbanization*, Sector Working Paper (Washington, D.C.: IBRD, June 1972), annex 1, table 4.

<sup>5</sup> Turnham and Jaeger, table V.1.

<sup>6</sup> According to Hans Singer, "Industrial employment in the modern sector is not increasing at the rate of 7-8 percent (the average increase in modern industrial production) but at the far lower rate of around 3 percent. Since total population also increases at a similar rate of close to 3 percent per annum, this means that there is practically no structural change in employment at all" ("Employment Problem in Developing Countries," *Manpower and Unemployment Research in Africa* 4, no. 1 [April 1971]: 31).

<sup>7</sup> "The kind of industrialization that has occurred in most Asian countries suggests that the employment-creating effects were weak. Most of the new industries appear to be no more than semiassembly plants which import a large part of their basic equipment and material inputs, producing consumer good which compete with traditional production, with very few linkages to tertiary sectors, etc." (Harry T. Oshima, "Labor-Force 'Explosion' and the Labor-intensive Sector in Asian Growth," *Economic Development and Cultural Change* 19, no. 4 [January 1971]: 164).

and for all, the crumbling tenements and flimsy shacks in which a majority of the urban population are herded together. Other urban services—such as water supply, sewage disposal, health care, primary education, and transportation—also lag behind even minimal needs.

Most disconcerting of all, occasional spurts in urban employment seem only to bring more migrants to the city. The phenomenon has not been studied with the care it deserves, but one suspects that, during most of the postwar-postindependence period, the elasticity of migration as a function of urban employment must have been greater than unity. If this relationship should continue to hold, as it seems likely to do, there is no chance whatever of “solving” the urban crisis in the developing countries, unless thoroughgoing and profound changes in policy are adopted and carried out. Some nonconventional thinking is beginning to appear in the specialized literature.<sup>8</sup> Given the relatively limited prospects of expanding agricultural employment, the starting point in each analysis is the question of how continued increases in labor force might be productively absorbed in the urban economy. The present study is concerned with this question.

In the following pages, we present a heuristic model of the urban labor market in developing countries. The model is intended as a beginning for empirical analysis. We believe that the model is generally applicable to countries with urban growth rates of roughly twice the rate of increase in national population, and whose larger cities have a “modern” sector that accounts for at least 15 percent of the urban work force. Implicit in the model are major implications for development policy. Before discussing these, one would normally wish to wait for a larger data base and empirical verification of the hypotheses advanced. The issues with which we are concerned, however, are of such magnitude and practical urgency, and the process of empirical validation is usually so drawn out, that we intend to outline major proposals even before all the evidence has been collected and analyzed.

### The Structure of Urban Employment

The urban economy in a developing country can be divided into three major employment sectors distinguished according to the organizational form of activities within each sector. In figure 1, these sectors are arranged

<sup>8</sup> Michael P. Todaro, “A Model of Labor Migration and Urban Unemployment in Less Developed Countries,” *American Economic Review* 59, no. 1 (March 1969): 138–48; Hans W. Singer, “A New Approach to the Problems of the Dual Society in Developing Countries,” *International Social Development Review*, no. 3 (1971), pp. 23–31; Ronald G. Ridker, *Employment in South Asia: Problems, Prospects, and Prescriptions*, Occasional Paper no. 1 (Washington, D.C.: Overseas Development Council, June 1971); Walter Galenson, ed., *Essays on Employment* (Geneva: ILO, 1971); Charles R. Frank, “Urban Unemployment and Economic Growth in Africa,” *Oxford Economic Papers* 20 (July 1968): 250–74; and International Labour Office, *Fiscal Measures for Employment Promotion in Developing Countries* (Geneva: ILO, 1972).



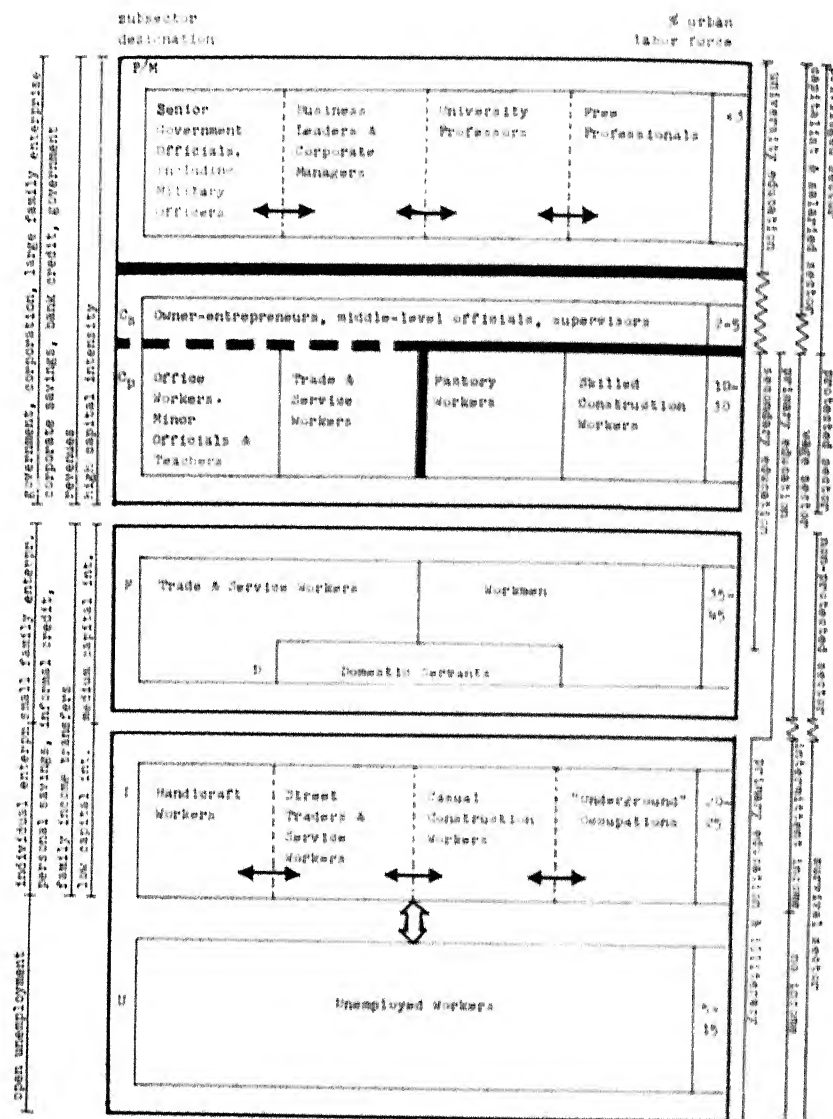


FIG. 1.—Structure of urban employment in a developing economy. Agricultural labor force residing in urban areas is excluded from this model. Unemployed workers belonging to the educated elite (P/M- and C<sub>a</sub>-sectors) are not formally included in this model, but see the text and fig. 3.

hierarchically in an ascending scale of labor productivity, economic power, and social status.<sup>9</sup>

<sup>9</sup> For reasons common to all scientific effort, our model represents a gross simplification of reality. In that lies both its virtue and its weakness. For example, we have had to ignore the fact that in some countries certain urban occupations are largely restricted to particular caste and ethnic groups. This introduces important modifications to our analysis. Nevertheless, we believe that the overall structural and dynamic features of the model continue to be a useful guide to empirical study.

Occupying the lowest rung, the individual-enterprise sector (I) includes the unemployed and self-employed workers that make up the "street economy" of a city. It is followed by a family-enterprise sector (F) consisting of workers in small trade and service establishments and industrial workshops having fewer than 50 employees and a low capital-to-labor ratio. Highest in rank is the corporate sector (C), which includes, in addition to workers in corporate enterprise properly speaking, those in family-run businesses that are larger than establishments in the F-sector, as well as the government bureaucracy, universities, and free professions.

If urban employment sectors are broken out in this way, a number of characteristics are revealed that correlate with the primary criterion of organizational structure. These are shown in the margins of figure 1. The I- and F-sectors, for instance, are for the most part self-financed, whereas the C-sector has direct access to government revenues, bank credit, and corporate profits. Partly as a result, capital intensity (and therefore labor productivity) rises progressively with each step in the hierarchy. (Some benchmark estimates of the amount of capital per

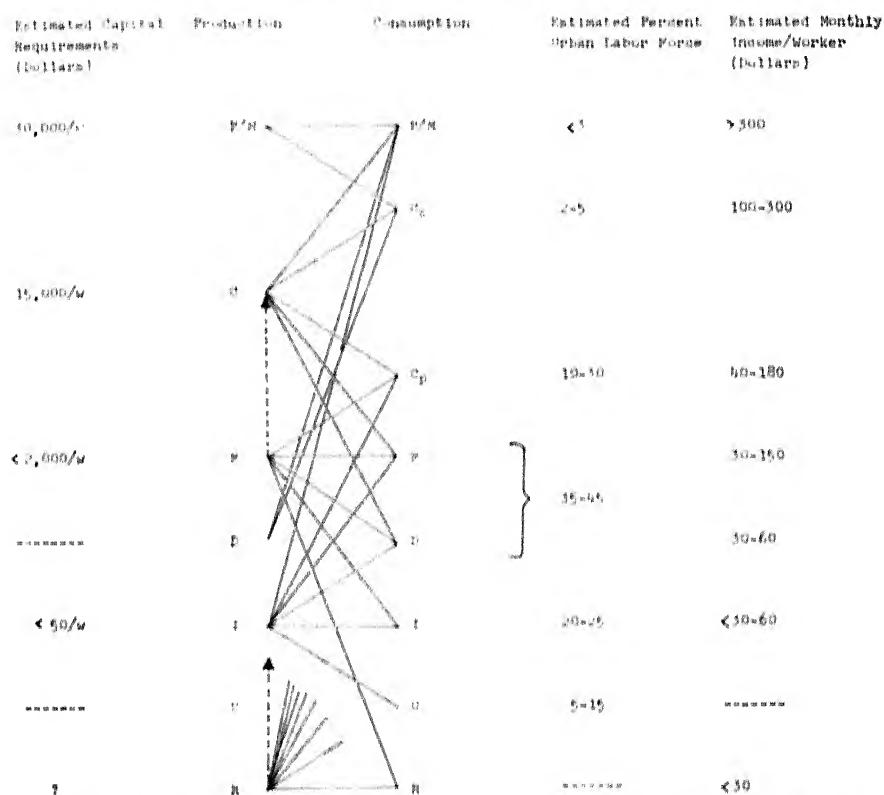


FIG. 2. Principal market relations among urban occupational and rural sectors. See fig. 1 for sector designations; R = rural sector.

worker are shown in fig. 2.) On the right-hand side of figure 1, additional characteristics of each sector relate to educational thresholds, sources of income, and degree of social and economic protection.

Summarizing this information, we can say that each of the three major employment sectors functions as a distinct subsystem of the urban economy. The functional linkages among these subsystems are not very well understood, though in some countries, notably in Mexico and Japan, the C- and F-sectors are interrelated in manufacturing through sub-contracting. In addition to the economic attributes already mentioned, each sector is characterized by social relations and rules of ethical conduct. The significance of this dimension, particularly the widespread social obligation of mutual aid and income sharing, will become more apparent in the subsequent discussion. Each sector and subsector will now be briefly described.

1. *The individual-enterprise sector.* The lowest-ranking subsector in our model of the urban economy includes unemployed workers (U) in the following three categories: first-time job seekers (the offspring primarily of urban residents and for the most part dependent adolescents), recent migrants to the city, and workers laid off from jobs in any of the subsectors above, up to and including corporate production (C<sub>p</sub>). Although they are not formally included in the model, white-collar unemployed should also be mentioned here. Members of this group may be considered downwardly mobile from professional-managerial (P/M) and C<sub>s</sub>-subsectors, and, even though they may never have worked, are disinclined to accept jobs below the level for which they think themselves qualified by reason of superior education and/or social standing (see fig. 3).<sup>10</sup>

New entrants into the labor force, mostly school leavers, are the most numerous group, comprising between 40 and 60 percent of the urban unemployed.<sup>11</sup> Because of their age and lack of work experience, young adolescents are at a competitive disadvantage with respect to recent migrants and may remain unemployed for considerable stretches of time.<sup>12</sup> (In some countries, even 10–12-year-olds may be counted among the labor force.) Migrants to the city tend to be the most qualified workers. They are often equal, if not superior, to the resident population in education,

<sup>10</sup> "Withholding of labour from the labour force is a common feature in India amongst the more educated, since it is felt, with some cause, that to take a job in a lower-level occupation preempts chances to improve one's occupational position later. Thus labour is withheld as long as it is economically possible until a job in a suitable occupational category is available" (Stewart F. Richards, "Geographic Mobility of Industrial Workers in India," in *The City as a Centre of Change in Asia*, ed. D. J. Dwyer [Hong Kong: Hong Kong University Press, 1972], p. 89).

<sup>11</sup> Ridker, p. 12.

<sup>12</sup> Unmarried dependents tend to have the highest unemployment rates, while heads of households have generally the lowest rate (see Turnham and Jaeger, table III-5). For Nigeria, see the interesting estimate prepared by P. Mueller and K. H. Zevering, "Employment Promotion through Rural Development: A Pilot Project in Western Nigeria," in Galenson, p. 94.

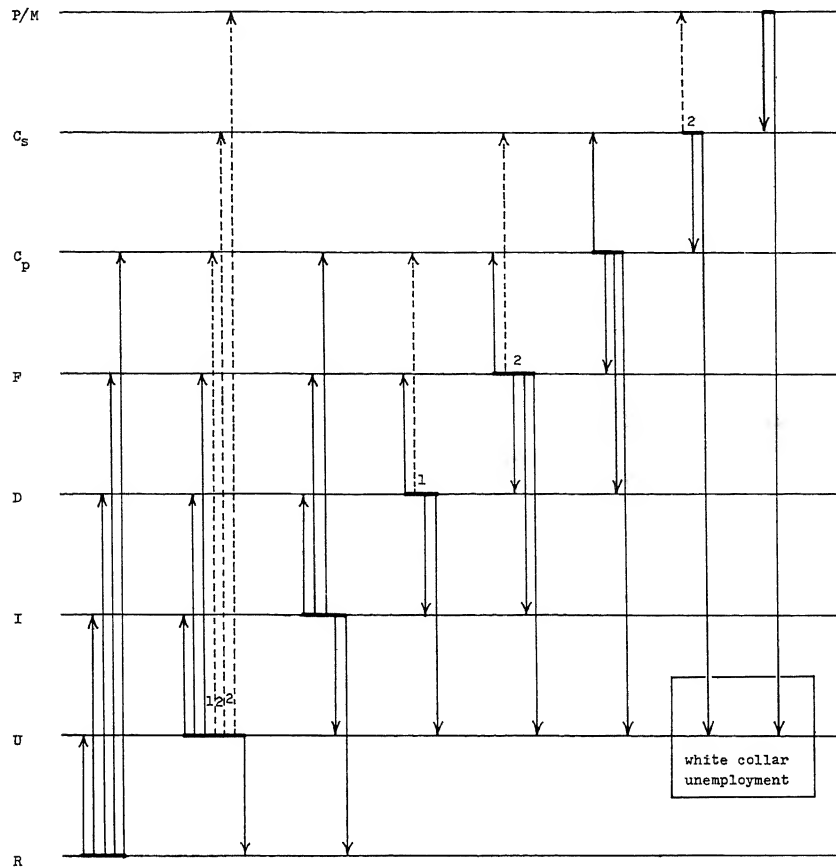


FIG. 3.—Intersectoral labor mobility. See fig. 1 for sector designation; R = rural sector; 1 = blue collar only; 2 = white collar only.

ability, and skills, and there is evidence that, after their arrival in the city, most migrants remain in open unemployment for only short periods of time.<sup>13</sup> Workers laid off from their jobs, comprising the third category, are likely to be either less skilled than those continued in employment or, alternatively, to have worked in the least protected and most volatile sectors of the urban economy (I and F). Some of them may eventually

<sup>13</sup> Most migrants find some work within a few days after arrival in the city. After about 6 months, migrants tend to have an employment distribution similar to that for the urban labor force as a whole (see, for example, Joan Nelson, "The Urban Poor: Disruption or Political Integration in Third World Cities," *World Politics* 22, no. 3 [April 1970]: 398; and John Stuart McDonald, "Migration and the Population of Ciudad Guayana," in Lloyd Rodwin and Associates, *Planning Urban Growth and Regional Development: The Experience of the Guayana Program in Venezuela* [Cambridge, Mass.: M.I.T. Press, 1969], p. 114).

return to the rural economy (R) for shorter or longer periods. The permanent abandonment of the city is rare, however, and involves primarily older people retiring from active life.

The next-ranking, or individual-enterprise subsector (I), may in practice be difficult to distinguish from open unemployment. Workers move readily from U to I and back again. Moreover, for each business transaction in the I-subsector, the time actually spent on the job may be quite long. Hawkers may pace the streets from sunup to sundown, completing only a few sales. Even so, it is necessary to distinguish this subsector because, in contrast to the destitute workers in U, its members do produce some income, however small, and contribute to family upkeep.

The I-subsector includes handicraft workers (seamstresses, embroidery makers, basket and mat makers, rope makers, silversmiths) working on their own account, street traders and service workers (peddlers, shoeshine boys, parking lot attendants, messengers, street entertainers, repairmen, gardeners, masseurs, food vendors, public letter writers), casual construction workers (carpenters, bricklayers, plumbers, electricians), and "underground" occupations (prostitutes, professional beggars, police spies, dope peddlers, pickpockets). There tends to be a good deal of lateral movement of workers from one category to another, with some workers having had experience in all categories and others actually holding down several jobs at a time. This pattern makes for an astonishingly versatile urban experience and may facilitate upward job mobility later on.<sup>14</sup>

Workers in the individual-enterprise subsector are typically in business for themselves. The situation of some of them may be so stable that they may justifiably be called self-employed. For others, however, such designation may be too dignified; they have to hustle jobs wherever they can. Their place of work is typically the street or the bazaar, but occasionally, as with handicraft workers, also the home.<sup>15</sup> Central city locations are of vital importance to these workers who, as opposed to squatters, generally occupy overcrowded and unhealthy quarters in the older sections of the city.<sup>16</sup>

<sup>14</sup> Hawkers have come to be recognized in Malaysia as an important sector of the urban economy and are being encouraged by their government through loans and the provision of mobile carts fitted for various hawking purposes. Hawking involves skills not easily learned. The hawker must know where the supplies are, the process of establishing credits or other arrangements to obtain the goods to sell, and the various procedures involved in transporting of goods from one place to another (including the payment of "tong"—bribes—to officials at the border of each district to assure quick approval of goods passage) (Chung-Tong Wu, personal communication).

<sup>15</sup> For a detailed description of the street or bazaar sector of the urban economy, see Clifford Geertz, *Peddlers and Princes: Social Development and Economic Change in Two Indonesian Towns* (Chicago: University of Chicago Press, 1963).

<sup>16</sup> An interesting classification of housing types for poor populations in Lima, Peru, is presented by Carlos Delgado, "Three Proposals regarding Accelerated Urbanization Problems in Metropolitan Areas: The Lima Case," in *Latin American Urban Policies and the Social Sciences*, ed. John Miller and Ralph A. Gakenheimer (Beverly Hills, Calif., and London: Sage Publications, 1971).

The market for I-workers is an intensely competitive one, and job specialization often reaches extreme forms with individuals competent in many different jobs. The available work is spread among the largest possible number of workers at the expense of income which is not only very low but frequently intermittent. Estimated income per worker in the I-subsector may range from less than \$30 per month (roughly equivalent in real terms to a rural wage) to \$60 (see fig. 2). If we take the upper figure as the urban subsistence minimum, many earning less than this amount necessarily fall burden to their kinsmen (or members of their village and/or tribal association), who are then obliged to share whatever surplus they may have, in cash or kind, with the less fortunate members of their own group.<sup>17</sup> Both the U- and I-subsectors, therefore, tend to subsist in a survival economy in which a downturn in urban production may force the least protected workers to return periodically to the countryside (see fig. 3). A significant recovery in the urban economy, on the other hand, is likely to bring additional job seekers into the marketplace and lead to a further fractionizing of the available work (see below).

A good many members of the U- and I-subsectors are officially invisible, and only the most imperfect, unreliable records are kept of their actual number. Most job classification systems break down when confronted with the great diversity of job experiences; the temporary, uncertain nature of the jobs themselves; and the horizontal movement among jobs of individuals in the U- and I-subsectors. Workers live in squattments that are frequently considered illegal and excluded from official city maps. Some may live on the street. When they operate without the necessary license, they are hunted down by the police. Where work permits are required for urban residence, their very presence in the city may be illegal.<sup>18</sup> Underground occupations, of course, are by their very nature "invisible" and may only be recorded on police records. A large proportion of the urban labor force, living within the interstices of urban society, has therefore no legal existence at all. This, surely, is one of the most incredible facts of city life in the developing countries.

2. *The family-enterprise sector.* This sector is distinguished from the I-subsector in that it (a) involves a higher degree of organization, (b) employs both wage and unpaid family workers (the latter working primarily for room and board), (c) has a fixed abode, and (d) operates with

<sup>17</sup> The subsistence minimum of \$60 is a purely hypothetical illustration and is intended to provide for the worker and the average number of per worker dependents not in the labor force. For the sake of comparison, the monthly family budget for minimum wage purposes in Mexico City in 1963 was valued at \$58.64 (see N. N. Franklin, "Minimum Living Standards in Latin America," in *Workers and Managers in Latin America*, ed. Stanley M. Davis and Louis Wolf Goodman [Lexington, Mass.: D. C. Heath & Co., 1972], p. 38).

<sup>18</sup> In Djakarta, for instance, the population residing illegally in the city has been estimated at several hundred thousand.

a larger amount of capital per worker. Here, again, statistics are most unreliable indicators of the numerical importance of this sector; in some countries, establishments with fewer than five or even 10 workers are not even counted in national censuses. But a rough guess suggests that small-scale family enterprise may employ between 35 and 45 percent of the urban labor force. Clearly, this is a key sector of the urban economy.<sup>19</sup>

The sector is roughly divided between (a) trade and service establishments and (b) manufacturing workshops. (Domestic servants are also included here as members of the household economy in the P/M- and C<sub>s</sub>-subsectors. In many respects, their situation resembles that of workers in the family-enterprise sector as a whole.) The former include small shop proprietors, salesmen in small businesses, garage mechanics, truck, bus, and taxi drivers, barbers, beauticians, and restaurant workers, among others; the latter include workers in various kinds of small-scale manufacturing, such as bakeries, rice mills, noodle shops, shoe and leatherware manufacture, tailoring, carpentry, and metal work. For the most part, they are engaged in the production of traditional commodities that are characterized by their appeal to a low-income mass market, a lack of standardization and quality control, and the use of indigenous raw materials. The pricing of F-sector commodities and services is exceedingly competitive; because they embody a high proportion of labor costs, the lower price limit (usually established in bargaining) is the survival capacity of the family group engaged in production. As a result, it may be difficult for workers in the F-sector to earn significantly more than the amounts required for family subsistence. Where a small surplus does accumulate as a result of competitive advantage and/or superior business acumen, it is often retained at the risk of antisocial conduct in which the obligation of sharing one's income with a wider group of claimants is at least partially avoided.

<sup>19</sup> In Kenya, "Small African businesses come in a wide variety of forms. There are small-scale manufacturers who utilize salvaged materials and who operate in extremely primitive quarters. There are the vehicle mechanics who, with several helpers, operate from vacant lots with very little by way of tools. There are the petty traders who dominate almost all traditional markets. On the other extreme, there are the manufacturing firms in Kenya Industrial Estates and a few relatively sophisticated commercial operators. . . . Total employment provided by all small-scale African business appears to be somewhat on the order of 480,000 or over 60 percent of the total non-agricultural employment. To put these numbers into perspective, it should be pointed out that these estimates of employment in small business largely do not find their way into official employment statistics; most of the enterprises are unregistered and, indeed, many are illegal in the sense that they do not satisfy statutory operating obligations (some of the activities themselves are illegal, e.g., the brewing of alcoholic beverages)" (*Report of the Working Party on Small Business Development* [Nairobi, June 1972], pp. 2-3). (These estimates include employment in what we have called the F- and I-sectors.) For a comparison with Asian countries, see Oshima, p. 165, and for the Middle East, United Nations Industrial Development Organization, *Small-Scale Industries in Arab Countries of the Middle East* (New York: United Nations, 1970).

Given its requirements for equipment and operating capital, this raises the question of how production in the F-sector is financed. The capital markets in which this sector operates lie outside the typical institutional sources of financing, such as banks and government loans. Although this seems to lead to a more realistic pricing of production factors than in the corporate sector with its heavy subsidies to capital, it is a handicap that renders capital accumulation in this sector quite precarious. Clans of related families may pool their resources to launch a business that may ultimately benefit some of their own members. Or a small firm may get started as a result of windfall gains, such as winning a lottery or the annual bonus that may be given to employees of large corporations. Additional income for investment may also be obtained from housing rentals. Speaking of the *bustees* (squatter settlements) in India, Colin Rosser writes:<sup>20</sup> "In Calcutta City alone, the 30,000 huts in the *bustees* are owned by 20,000 hut owners... and rooms in them are let out to the tenant population of over 700,000." Lisa Peattie reports on a similar phenomenon in Venezuela's Ciudad Guayana.<sup>21</sup> Occasionally, as in the case of family production units selling directly to firms in the corporate sector, the latter may advance both cash and raw materials to assure themselves of a continued source of supply. Finally, loans may be obtained from private money lenders, even though the interest charges on such loans are usually exorbitant. But in no country is regular bank credit (or government financial assistance) to the F-sector available in anywhere near the amounts required.

3. *The corporate sector.* The corporate-production sector ( $C_p$ ) ranks immediately above the F-sector, but, qualitatively it is very different. Its workers are members of bureaucratic organizations; they operate with capital equipment that may be 10 times or more the value used in the F-sector; and the conditions of their employment are legally protected. The legal protection they enjoy is related to the large, impersonal organization to which they belong. Labor legislation is chiefly aimed at this sector; at least, it is enforced there more than elsewhere and places  $C_p$ -workers in a contractual (rather than personal) relationship to their employers. Such legislation varies among countries but may include regulations concerning the employment of child labor, regulated working hours, job security, minimum wages, paid vacations and sick leave, special food supplements, preferential claims on housing, etc. Workers in the  $C_p$ -subsector are, indeed, a favored group. Some of the advantages accruing to them are the result of a militant unionism that is concentrated here;<sup>22</sup>

<sup>20</sup> Colin Rosser, "Housing and Urban Change in Calcutta," in Dwyer, p. 186.

<sup>21</sup> Lisa Peattie, "Social Mobility and Economic Development," in Rodwin and Associates, p. 405.

<sup>22</sup> Peter Kilby, "Industrial Relations and Wage Determination: Failure of the Anglo-Saxon Model," *Journal of Developing Areas* 1, no. 4 (July 1967): 489-520.



more often, their advantages stem from social legislation passed in imitation of advanced industrial societies.

In most countries, important social distinctions are made between white- and blue-collar workers.<sup>23</sup> Such distinctions make little sense in any of the lower-ranking sectors and subsectors; here they serve as a basis for granting special privileges to the salaried class, such as shorter working hours, longer vacations, maternity leave, access to specially designated stores where goods may be purchased at subsidized prices, preferential housing loans, and so forth. As a result, lateral mobility between white- and blue-collar occupations is relatively low within the C<sub>p</sub>-subsector, and school leavers, particularly those with a secondary education, may remain unemployed for relatively long periods in hopes of gaining a salaried position. As in bureaucratic organizations generally, skill levels are carefully differentiated in this subsector. At the lower end of the scale, generally applied to women and children, the average wage may be below subsistence, but at the highest levels, in such white-collar jobs as section chiefs in government ministries or senior clerks in private business, a salary of up to three times subsistence may be paid.

As the principal beneficiary of development policy, corporate manufacturing is also the most productive source of work in the urban economy. Gains in excess of 7 percent in industrial output are no longer unusual, but increases in employment are generally only half as much. Indeed, on a net basis, and accounting for the jobs eliminated in the F-sector by the expansion of corporate enterprise, employment changes for the industrial sector as a whole may be negative.<sup>24</sup> The capacity of the C<sub>p</sub>-subsector for absorbing employment is therefore exceedingly restricted, especially in view of its small initial base. Most of its gains will be in white-collar and, more specifically, in government employment. Except for seasonal variations, construction tends to be a fairly constant percentage of employment over time.<sup>25</sup> Only 10–30 percent of the urban labor force are employed in this subsector, are casual laborers in their majority, and are more properly counted as belonging to the I-subsector of the economy.

<sup>23</sup> "The most basic distinction made in the stratification system in all Mexican industrial firms is between *empleados* and *obreros*" (Stanley M. Davis, "*Empleados and Obreros*," in Davis and Goodman, p. 31).

<sup>24</sup> As Stanback and Knight have pointed out, net measures of employment changes seriously understate the extent of change which has taken place. "For the period 1950 to 1960 in U.S. metropolitan areas, about 23 jobs were eliminated for every 100 jobs created" (Thomas M. Stanback and Richard V. Knight, *The Metropolitan Economy* [New York: Columbia University Press, 1970], p. 2).

<sup>25</sup> For recent years in nine underdeveloped countries, the annual employment growth rate in the construction sector was only 1.3 percent. In a smaller number of more developed countries, the growth rate was 5.1 percent. On the other hand, the construction sector accounted for only 3.9 percent in the first and 6.7 percent in the second group of countries (see W. Paul Strassman, "Construction Productivity and Employment in Developing Countries," in Galenson, p. 146, table 1).

Although the corporate supervisory subsector ( $C_s$ ) is closely integrated with production on a functional basis, it is sharply separated socially, a fact that is indicated by the heavy horizontal line in figure 1. The basis for social differentiation is primarily education, a completed secondary or university education being generally required to rise into supervisory corporate positions. This tiny employment subsector, accounting for only 2–5 percent of the urban labor force, nevertheless constitutes the backbone of the nascent middle class, though minor officials and clerical workers in the  $C_p$ -subsector may likewise aspire to preferred social positions and adopt middle-class life styles to the extent permitted by their incomes.

Separated from the  $C_s$ -subsector is the city's professional-managerial elite (P/M). Accounting for less than 3 percent of the labor force, this employment subsector of senior government officials and military officers, business leaders, corporate managers, university professors, and free professionals is set apart from those below it by traditional family status, wealth, and education. As in the lower-ranking I-subsector, employment in P/M is characterized by a good deal of lateral movement, and individuals may frequently hold two or three jobs across several occupational groupings. For each one of these, remuneration may be on the order of \$300 a month (or five times the assumed subsistence minimum). By cumulating jobs, however, and by receiving additional income from real estate investments, business profits, and bribes, total income may be very much higher than this.<sup>26</sup> The urban elite is consequently the only subsector of the urban economy capable of accumulating a sizable income surplus. It is protected against excessive sharing of incomes (or wealth) by the impermeable social barriers which separate it from lower-ranking employment subsectors. Much of this surplus, however, will be channeled into conspicuous consumption such as villas, automobiles, servants, jewelry, travel, private education, and art objects, and into "safe" but profitable investments such as real estate.<sup>27</sup>

<sup>26</sup> Income distribution estimates for developing countries suggest that the top 10 percent of the population receives between one-third and one-half of total income (see Turnham and Jaeger, table IV-1; also, Economic Commission for Latin America, "Some Regional Development Problems in Latin America Linked to Metropolitanization," *Economic Bulletin for Latin America*, vol. 16, no. 2 [1971], tables 20–22).

<sup>27</sup> "The successful Filipino entrepreneur, in company with other successful Filipino businessmen, has created a suburban and upper-class style of life: . . . a modern air-conditioned office; membership and frequently elective office in the Lions' Club or Rotary, a golf club, and one or more business associations; a large and very modern home in one of the suburbs of Manila, air-conditioned, and often with a swimming pool; a summer home in the mountain city of Baguio; a prestige school in the Philippines and/or education in the United States for his children; frequent travel abroad for the entrepreneur and his family; two or more automobiles and a great variety of fine clothes for the entrepreneur and his family. Occasionally . . . it includes a *querida* or mistress" (J. Carroll, *The Filipino Manufacturing Entrepreneur* [Ithaca, N.Y.: Cornell University Press, 1965], p. 134, quoted in Leandro A. Viloria, "The Manileños," in Dwyer, p. 27).

**The Dynamics of Urban Employment: Three Hypotheses**

Based on the foregoing model, three hypotheses may be advanced, each of which, if proved correct, will have tremendous practical significance for urban development policy.

1. *Equilibrium unemployment.* The percentage of the labor force in the individual enterprise sector ( $U + I$ ) tends to remain fairly stable over a wide range of variable conditions. To a certain extent, it may be regarded as a "holding" sector for new entrants into the labor force, including rural migrants, with workers queueing up to wait for more steady employment in the family enterprise and corporate sectors ( $C_p$ ). Some I-employment may itself be steady and permanent, however, whereas certain F-jobs, to the extent they are below subsistence, may be regarded as temporary, "holding" occupations.

The maintenance of a sizable I-sector (including up to 15 percent of open unemployment) requires explanation. According to economic theory, one would expect large-scale unemployment and "underemployment" to depress wages in high-ranking sectors until all but fractional unemployment is absorbed. But for a variety of reasons, this does not occur, and the I-sector continues to provide an entry into the urban economy for between 25 and 40 percent of the urban labor force.

To explain this apparent anomaly, it may be suggested that the rate of entry into I is primarily controlled by (1) expectations of finding steady employment at subsistence levels or better<sup>28</sup> and (2) the total surplus over subsistence available in the urban economy, principally from F- and  $C_p$ -subsectors. The expectational calculus in this model acts as the principal encouragement to urban migration; the greater the expectation of finding a good job, the larger will be the volume of migration (holding constant for subsistence surplus). On the other hand, the available income surplus establishes an upper limit to the absorption of labor force into the urban economy. Existing I-sector employment (and unemployment) will therefore be in equilibrium (given a "normal" level of expectations for finding employment at or above subsistence) subject to the constraint of the available surplus necessary to maintain the labor force in I. A significant improvement in the urban job situation will tend to raise expectations, accelerating the inflow of migrants into the city above the actual increase in the number of jobs. (The employment elasticity of migration is assumed to be greater than one.) Any slack in the urban economy will, therefore, be quickly eliminated, with expectations (and migration) returning to their original levels. Similarly, as more people move to the city, any existing increase in the available surplus will be used up (given the incapacity of the urban economy to absorb all newcomers into employment

<sup>28</sup> This hypothesis is taken, in modified form, from Todaro. It should be noted that the estimated urban subsistence minimum has a nominal value of more than twice the average rural income.

at subsistence or above). This will tend to depress living levels to just below the acceptable subsistence minimum, rendering migration less attractive until such time as employment and unemployment equilibrium in the I-sector are restored.

This hypothesis assumes, quite reasonably, we believe, that the communication system linking cities to distant villages and signaling the job situation in the urban economy functions with considerable speed and accuracy, though it will obviously work more efficiently along the "ridges" of high population potential (reflecting both the volume and separating distances of population) than in the intervening "valleys" of more isolated and dispersed populations. On the other hand, we have excluded from consideration such other, noneconomic incentives as perceived greater access to services (education, health, entertainment) in the urban environment which, in the developing countries, are frequently claimed to be significant determinants of rural to urban migration.

2. *Pressure to subsistence.* A worker can expect to survive in the urban economy only to the extent that he earns a subsistence minimum which we have somewhat arbitrarily established at about \$60 a month.<sup>29</sup> Social pressures for income sharing are often very great for kinsmen, hometown, provincial, caste, and tribal members, particularly in the F- and C<sub>p</sub>-subsectors.<sup>30</sup> Thus, a clerk employed in the C<sub>p</sub>-subsector earning \$100 a month can support both himself and two additional I-workers (say, his wife and his wife's cousin), each of whom contributes \$40 to the family's monthly income. Alternatively, if both husband and wife have full-time jobs in the C<sub>p</sub>-subsector, with a combined monthly income of \$200, they can support not only themselves (\$120) but also one unemployed kinsman (say, the husband's brother, at \$60) and the wife's cousin, earning an average monthly income of \$20. A third scenario involves a family workshop (F)

<sup>29</sup> The subsistence minimum is not necessarily a physical minimum but is related to accepted standards of consumption. Under certain conditions, it may be lowered down to a level of actual starvation.

<sup>30</sup> This phenomenon is almost universal (see, for example, Kenneth Little, *West Africa Urbanization* [London: Cambridge University Press, 1965], pp. 47-65; William Mangin, "The Role of Regional Associations in Adaptation of Rural Migrants to Cities in Peru," in *Contemporary Cultures and Societies of Latin America*, ed. Dwight Heath and Richard Adams [New York: Random House, 1965], pp. 311-24; Lisa Peattie, *The View from the Barrio* [Ann Arbor: University of Michigan Press, 1968]; Madelaine Trebous, *Migration and Development: The Case of Algeria* [Paris: Development Centre of the OECD, 1970]; Ann Wee, "Some Social Implications of Rehousing Programmes in Singapore," in Dwyer, pp. 211-24; and Mary R. Hollin- steiner, "Becoming an Urbanite: The Neighborhood as a Learning Environment," in Dwyer, p. 32). Two brief quotes from the studies above make the point very succinctly: "In La Laja, with one out of three adult males out of a job, and nearly half of the population under thirteen, one person out of six had any regular income. The other five-sixths were apparently living—if precariously—off the employed segment" (Peattie, *View from the Barrio*, p. 45); "Individual savings are minute since every person who has a job has to support six, eight, twelve, or more dependents" (Trebous, p. 27).

with a monthly net income of \$600. This enterprise employs eight workers in addition to the owner: a foreman who is paid \$100, a foreman's assistant at \$80, three workers to whom the owner has no direct family obligation and who together receive \$120 (they will have to find additional support among their own kin), and three unpaid family workers whose full support costs the owner an estimated \$180. This amounts to a total imputed wage bill of \$480, leaving the owner himself with twice the subsistence wage. A portion of this remainder, however, may have to be used to help other needy relatives.

These examples illustrate one typical form of income sharing. Equally important, however, are income transfers from urban migrant workers to their places of origin. In the aggregate, these may account for a sizable proportion of rural incomes. Whatever the form of income sharing that is used, however—and different forms may be in use at the same time—the result will be a relatively flat income distribution across all urban subsectors, from U to C<sub>p</sub>, at close to the subsistence minimum of \$60.<sup>31</sup>

If, for any given volume of work, the ultimate limit to urban size is at subsistence levels, urban services—including housing—will always be stretched to the maximum. The more housing is built, especially for lower-income groups, the more income in kind is available for sharing, and the higher will be the rate of immigration from rural areas (subject to the constraint of job expectations). Refusal to share available surplus will, of course, be rewarded with a disposable income (and level of material well-being) above subsistence. But for many urban dwellers, each of whom may some day be himself in need of help, refusal is not an available option. Family (and tribal) obligations must be honored. In developing countries, the moral economy rewards shared poverty.<sup>32</sup>

3. *Proletarianization.* Shared poverty is the common experience of population employed in the I- and F-sectors and also, to some extent, in the less protected niches of the C<sub>p</sub>-subsector. Any income gains in these sectors will be quickly dissipated among new arrivals to the city. Incomes remain above subsistence primarily in the C<sub>s</sub>- and P/M-subsectors. The income gap between the urban poor—comprising the vast majority of the

<sup>31</sup> Available data on income distribution fail to reflect this tendency fully, since they refer exclusively to income earners rather than income users. Nevertheless, even on that limited basis, it appears that for over at least 70 percent of all income earners in Latin America, differences in income received are relatively small. Income inequalities arise chiefly from the rather steep climbs of income beyond the seventieth percentile (see Economic Commission for Latin America, *Income Distribution in Latin America* [New York: United Nations, 1971], pp. 18–20).

<sup>32</sup> Some successful business ventures in this sector do exist, however, which either have few demands placed upon their resources or are able to meet numerous demands without suffering excessive losses.

population—and the privileged sector will therefore increase over time.<sup>33</sup> Moreover, the continued expansion of production in the corporate sector, especially manufacturing, will tend to drive out “inferior” goods and services provided in the F-sector with which it is, in part, competitive. Because of higher labor productivity, job increases in the corporate sector will not only be less in proportion to production gains but will often destroy a large number of jobs in similar industries within the F-sector.<sup>34</sup> This situation, coupled with the limited capacity of the rural economy to absorb additional labor, means that the expansion of “modern” industry ( $C_p$ ) will drive labor from affected regular employment in small-scale enterprise into the I-sector. Theoretically, one would expect that downward mobility of labor from F to I will be compensated by a commensurate decline in migration from rural areas. In fact, however, rising expectations of employment in the  $C_p$ -subsector will cause urban migration to be at least maintained at prevailing levels, if not actually increased. But this is likely to be accomplished only through a lowering of the urban subsistence standard. Accelerated industrialization in the corporate sector will, therefore, be accompanied by increased unemployment, a lowering of labor productivity in the aggregate, and the growing impoverishment of the population in both the I- and F-sectors of the urban economy.<sup>35</sup> This process, of course, is bounded, at the lower limit, by the absolute subsistence income at which an urban population can survive.

#### Market Relations and Labor Mobility in the Urban Economy

Additional aspects of the urban employment model are summarized in figures 2 and 3. The diagram of principal market relations (fig. 2) suggests the following conclusions.

<sup>33</sup> This tendency is exacerbated by inflation, which in the developing countries is typically in excess of 10 percent a year (see United Nations, *1967 Report on the World Social Situation* [New York: Department of Economic and Social Affairs, 1969], p. 74, table 8). (This report is a generally excellent source of information on the social conditions of employment throughout the developing countries.)

<sup>34</sup> The best and most graphic demonstration of this tendency comes from Keith Marsden, “Progressive Technologies for Developing Countries,” in Galenson, pp. 113–40. The following case is particularly illuminating: “One country imported two plastic injection-moulding machines costing \$100,000 with moulds. Working three shifts and with a total labour force of forty workers they produced 1.5 million pairs of plastic sandals and shoes a year. At \$2 a pair these were better value (longer life) than cheap leather footwear at the same price. Thus, 5000 artisan shoemakers lost their livelihood; this, in turn, reduced the markets for the suppliers and makers of leather, hand tools, cotton thread, tacks, glues, wax and polish, eyelets, fabric linings, laces, wooden lasts and carton boxes, none of which was required for plastic footwear. As all the machinery and the material (PVC) for the plastic footwear had to be imported, while the leather footwear was based largely on indigenous materials and industries, the net result was a decline in both employment and real income within the country” (p. 118).

<sup>35</sup> Much of the increased income from expanding production in the  $C_p$ -subsector will be drained off in the form of luxury consumption by the P/M-subsector and will only marginally affect domestic production.

1. The individual-enterprise subsector sells directly to all other subsectors in the urban economy. This helps explain the relative size of this subsector (20–25 percent of urban labor force). At the same time, job specialization is carried to extremes, the quality of workmanship is uncertain, business competition is fierce, and prices are depressed. This situation compels workers to establish personal, quasifamilistic relations with their clients, in a desperate effort to assure themselves of a small corner of the market. Business relations in this sector (and to a certain extent in the F-sector as well) tend therefore to conform to social as much as to economic rules of conduct, except that social rules tend to be suspended in the case of transient strangers who will then have to bear the full brunt of economic rationality.

2. The markets for F- and C-sectors are partially overlapping. For reasons of design, quality, and prestige (and sometimes even price), the C-sector will tend to dominate in most situations where the two sectors compete (e.g., supermarkets versus grocery stores) and give rise to the process of progressive proletarianization discussed in hypothesis 3 above.

3. In the case of the P/M-subsector, market relations apply—strictly speaking—only to university professors and free professionals. Clients for these occupational groupings are principally from families within the P/M- and C<sub>s</sub>-subsectors, and their number is consequently quite restricted. Given the small number of potential clients, even though their individual incomes may be high, the intensity of competition here is similar to that encountered in the I-subsector. For this reason, a good many of those active in the professions or university teaching are earning incomes below their station and are able to maintain an appropriate living standard only by accumulating several jobs at a time. Still others, though professionally qualified, are unable to establish themselves in practice at all and either remain unemployed (and a continued economic burden to their families) or are forced to accept employment in the lower rungs of the corporate sector. This is especially true for journalists, architects, lawyers, economists, and sociologists.

Economic insecurity at the top of the occupational pyramid is partly counteracted by the ability of the elite to defend itself politically through legislation that is favorable to itself. A guildlike system may evolve, restricting the entry of new members into the professions while laying exclusive claims to certain types of work. In many South American countries, for example, major buildings must, by law, be architect-designed. The medical professions everywhere (except under socialism) resists the creation of paramedical occupations. A further barrier is the high cost of a professional education which prevents those with less than elite status from acquiring the necessary credentials to practice.

The diagram of intersectoral labor mobility (fig. 3) shows some of the mobility paths discussed throughout this paper. (It should be noted that apparent “advancement” from D to F or “decline” from F to D are, in



fact, lateral movements within the same sector.) Few studies have been carried out, so that the hypothesis implicit in figure 3 must remain highly tentative. On the whole, however, the occupational mobility of individual workers is probably less significant than the mobility of sons with respect to their fathers' occupations. As T. G. McGee observes, "The basic pattern is one of occupational stability."<sup>36</sup> With these provisos, the following tentative conclusions may be drawn.

1. Individual mobility paths conform to a loose hierarchical structure. Although some subsectors may be skipped, top occupations are generally reached only from "superior" subsectors. The rise to the F-sector, however, can be accomplished with roughly equal probabilities from any of the subsectors below it. Downward mobility, on the other hand, particularly to U, is possible from any employment subsector without regard to hierarchical position.

2. Four major reasons may be advanced to account for downward job mobility: (a) the destruction of jobs in F by an increase of competitive production in C<sub>p</sub>; (b) overtraining in relation to available job openings, particularly at C<sub>s</sub> and P/M levels and in the white-collar groupings of the C<sub>p</sub>-subsector; (c) prolonged unemployment (or below-subsistence employment) coupled with inability or unwillingness of kinfolk or mutual aid societies to continue maintenance support (in this event, return migration may occur to rural areas); and (d) incompatibility of work discipline in the C<sub>p</sub>-subsector with personal inclinations. This may also be expressed in part as a preference for the "warm" climate of personal relationships which tends to prevail in F-sector employment.

3. Upward job mobility may require: (a) social contacts: a heavy investment of time in maintaining and strengthening kinship, tribal, friendship, and patronage relations is generally required to advance up the occupational ladder—the absence of such contacts will almost certainly foreclose most mobility channels; (b) education: credentialism is

<sup>36</sup> See, for example, José Luis Reyna, "Occupational Mobility: The Mexican Case," in Davis and Goodman, chap. 18. According to Mexican survey data for 1963 comparing sons' occupations with those of their fathers, 70 percent of the sample was nonmobile, 20 percent upwardly mobile, and 10 percent downwardly mobile. This is consistent with McGee's finding for Malays that "almost two-thirds of the post-independence migrants have experienced no interoccupational mobility in their shift to Kuala Lumpur. . . . The basic pattern is one of occupational stability" (T. G. McGee, *The Urbanization in the Third World* [London: G. Bell, 1972], pp. 167–68). A generally similar finding for a sample of Western Ibo migrants in Ibadan is reported by Daniel Isika in "The Employment of Mid-Western Ibo Migrants in the City of Ibadan" (B. A. thesis, Department of Geography, University of Ibadan, 1969), pp. 62–63 ff. Several reasons may be suggested to account for these results: (1) low degree of wage differentiation across most of the available spectrum of jobs, (2) a high incidence of credentialism, particularly in the higher occupational strata, (3) intense job competition for higher-level positions, and (4) a view that personal achievement fails to pay off in terms of upward job mobility and that advancement is primarily a matter of luck.



especially important for reaching white-collar occupations in the corporate sector,<sup>37</sup> since functional illiterates tend to be concentrated in the U-, I-, and D-subsectors (see fig. 1); and (c) capital: for most workers, little capital is needed to achieve upward occupational mobility. Participation in the I-subsector calls for only modest amounts of equipment, such as trade tools; larger amounts will be needed to start a business in the F-sector. Beyond the F-sector, of course, the picture changes, and the capital requirements for establishing a private professional practice turn out to be prohibitive for all but a small minority of well-situated individuals.

### Major Policy Implications

Our analysis has been couched in terms of a concept of labor absorption. But in reality, we have been discussing a structural problem of pervasive urban poverty. We have seen how, under present conditions in the developing countries, the economic system works to keep the majority of urban workers close to a minimum of subsistence. We might have pressed our analysis even further. We might have shown, for instance, how under accelerated inflation, a condition that seems endemic to the development process in capitalist societies, the corporate sector is able to improve its income position relative to that of population in the F- and I-sectors that lack the organized political strength successfully to defend their income position. A look at the immediate future suggests a further deterioration of a situation that has already reached crisis proportions in a number of countries. Relatively high wage costs in the corporate sector (together with subsidized costs of capital) tend to encourage further capital intensification and so to reduce the labor absorption capacity of this sector. In view of urban growth rates throughout many of the developing countries that point to a doubling of urban populations in less than 20 years, we can confidently expect a dramatic worsening of the urban employment situation. For the developing countries as a whole, Turnham and Jaeger project hypothetical open unemployment rates of between 15 and 20 percent by 1980, relative to an urban population 50 percent greater than in 1970.<sup>38</sup> But this increase in open unemployment is possible only if we also assume a gradual reduction in urban subsistence levels, say, from \$60 to less than \$50 per month per worker. This belt tightening could, in principle, continue to the point of mass starvation in cities. Although the rate of rural to urban migration may diminish with a decline in the urban subsistence level, a growing percentage of urban growth is expected to derive from natural increase. A small reduction in the rate of urban-bound migration will therefore no longer suffice to forestall the ultimate crisis.

<sup>37</sup> Credentialism in relation to migration is discussed in Jorge Balán, "Migrant-Native Socioeconomic Differences in Latin American Cities: A Structural Analysis," *Latin American Research Review* 4, no. 1 (Spring 1969): 3-29.

<sup>38</sup> Turnham and Jaeger, pp. 114-20.

These trends must be seen in a perspective view of the total economy. If the lower two-thirds of an expanding urban population are getting progressively poorer, the upper one-third and, more specifically, the less than 3 percent of the population who derive their incomes from the P/M-subsector, will be the principal beneficiaries of continued economic growth. In a situation such as this, marked by extreme and steadily increasing forms of inequality, governments have essentially two options: either to side with those who wield the instruments of economic power and use repressive measures against the poor and their advocates in order to maintain political stability, or to assume the active leadership of the mass of the population, both rural and urban, and to devise an economic system that will achieve continued economic expansion together with increased equality in the distribution of the product.

What might be the characteristics of such a system? Although we are able to give here only the barest of outlines, we are persuaded that any policies falling substantially short of those identified will at best provide for only a temporary improvement in living conditions and, at worst, produce no measurable results whatever.

1. *Development objectives: from maximizing growth in GNP to maximizing human potential.* Ever since Dudley Seers delivered his now-famous address to the Eleventh World Conference of the Society for International Development in 1969, an exclusive preoccupation with economic growth as the principal and overriding objective of development has come under serious and sustained criticism.<sup>39</sup> The argument has been chiefly based, however, on a disappointment with the results of conventional growth policies and particularly with their employment and distributional effects. The fact that in a GNP-oriented development approach the human element appears as only one among several factors of production, in principle substitutable for one another, has not received the attention it deserves.

Although economic welfare and human happiness are always cited as final ends of economic activity, the meaning of welfare and happiness is interpreted in the context of the utilitarian framework of Western economics, in which consumer satisfactions appear as the only appropriate measure. In the means-ends framework of economic analysis, therefore, the postponement of consumer satisfactions now for greater satisfactions later is taken to be axiomatic. Inevitably, therefore, the human contributions to development have to be regarded as instrumental to the ultimate consummatory act.

Our analysis, on the other hand, suggests that the postponement of consumer satisfactions demanded by classical economics might, indeed, in

<sup>39</sup> Dudley Seers, "The Meaning of Development," reprint (New York: Agricultural Development Council, September 1970), pp. 1-11.

the case of the developing countries and for the great majority of their populations, be a permanent one. If this is so, the development models of classical economics must be discarded, consumer satisfactions are no longer the arbiter of welfare and human happiness, and the postponement of individual gratifications becomes an irrelevant category.

An alternative model would place the maximization of human potentials as the guiding objective, with growth in GNP appearing as a by-product of a process pointed toward this end. By this formulation we simply mean that a country's development is synonymous with the full realization of the creative energies of its people.

A country that would adopt such an objective would no longer tolerate the calculated exploitation and growing impoverishment of its work force for the sake of a dominant minority or allow the profligate, squanderous treatment of its most precious asset. The implications of this basic change in orientation concerning the substance of development are spelled out in the following paragraphs.

2. *Economic organization: from a system based on inequality to one of greater equality and social justice.* The capitalist system of production has erected inequality as its basic organizing concept. This fact is frequently glossed over in economic theory with its explicit bias toward equilibrium and is forgotten in the Western democracies, where enlightened governments have created a welfare system that systematically transfers income from the rich to the poor. (In the United States, the settlement frontier first and, more recently, the ever-expanding technological frontier have served as partial substitutes for the welfare state.) But the widespread imitation of welfare legislation in the developing countries has served only to create enclaves of privilege (see fig. 1). It has not weakened the basic condition of inequality; on the contrary, inequalities have been exacerbated.

In country after country, those in control of the instruments of economic power have consistently perverted one of the essential purposes of government—to redress the major imbalances in the society—by using the power of public office to further entrench their privileged position. Because such practice generates widespread antagonism and hostility, repressive measures tend to be applied against a mounting opposition. Fascist regimes in which the coercive powers of the State are jointly used with private economic power to maintain and widen existing inequalities abound in the developing world: Brazil, South Korea, Taiwan, the Philippines, Pakistan, Turkey, Greece, Spain, Senegal, and Rhodesia are only some of the better-known instances of fascism among the poor.

An alternative principle of organization, stressing equality, is that of communal organization in which the irreducible unit of society is no longer the individual but the organized commune. The commune is built on the premise of worker self-management and on the further premise of

distribution according to need. Production communes, which would remain small enough to permit face-to-face contacts among their members, may be brought together into assemblies of any size according to the requirements of the production process through a system of worker representation.

This is not the place to enter into a detailed description of this alternate model. Indeed, the concept of equality admits of a number of variant organizational forms. Ultimately, it is not the specific form of social organization that matters, but the commitment to the underlying ideas. The point we wish to emphasize is the need to evolve an ideology grounded in the principles of social justice and equality that is also suited to national conditions, culture, and historical circumstance. Given such a commitment and in due course, the appropriate form of social organization will surely be discovered.

3. *External relations: from foreign dependence to greater national autonomy and self-reliance.* Dunn, Friedmann, and others have identified development with a learning society.<sup>40</sup> But a learning society is possible only where appropriate theory—drawn from a detailed examination of the environment—is applied and tested in practice. The multiple strands of external dependence—economic, technological, and cultural—which link developing societies to the metropolitan powers of Western Europe and the United States militate against an effective learning process of development. Development in the form of salient ideas, capital, technical know-how, cultural artifacts, and socioeconomic organization is to a large extent imported and encapsulated in “modernizing” enclaves within the principal cities of the host society. Among these foreign imports are the ideas that development is synonymous with maximizing growth of GNP, and that the only viable form of economic organization is one that rests on inequality.

If a country is to learn to deal forthrightly with its problems of accelerated urbanization, these ideas must be rejected. Development focused on human potentials must increasingly be based on a sense of national autonomy and self-reliance. Only the exercise of autonomous choice is capable of accomplishing the shifts in economic policy pertaining to industrial organization, regional development, and population that are necessary for the eradication of poverty and exploitation. In line with a national policy of self-reliance, domestic capital formation, based on human labor and ingenuity, must to a large extent replace the importation of foreign capital; the inventiveness of scientific and production workers must be encouraged and rewarded; and national ideologies, grown on

<sup>40</sup> Edgar S. Dunn, Jr., *Economic and Social Development: A Process of Social Learning* (Baltimore: Johns Hopkins Press, 1973); and John Friedmann, *Retracking America: A Theory of Transactive Planning* (Garden City, N.Y.: Anchor Press/Doubleday, 1973).

native soil, must be raised to become dominant features of the cultural landscape. All of this argues against the massive financial and technical assistance to developing countries that has been promoted as a keystone of Western liberal thought since the end of World War II.<sup>41</sup>

4. *Industrial organization: from import substitution to an explicit policy of industrial dualism.* None of the policies enumerated above would, of themselves, assist with the productive employment of a rapidly expanding urban labor force. They require a set of auxiliary policies which address themselves specifically to this issue.

The corporate manufacturing sector in developing countries, and the business and banking interest allied to it, are primarily the result of substituting domestic production for the importation of commodities whose consumption is largely restricted to the upper income groups of the urban economy. Given the propensity of this import-substitution sector to greater capital intensity and its inability to employ more than a small fraction of additions to the urban labor force—indeed, given its tendency to destroy existing employment in lines of similar production within a less efficient F-sector—future employment policies should seek to expand the production of small-scale industries while restricting corporate production to capital equipment, commodities for export, and items not conveniently produced elsewhere. In this way, the precarious position of the F-sector would be protected against corporate competition, much as the present-day corporate sector obtains protection from foreign competition through various forms of import restriction.

The principal need of a revitalized F-sector is access to credit for both investment and operating capital. Government credit needs to be channeled to this sector in increasing amounts. Other significant problems include a lack of supporting trade organizations to defend the interests of small producers, lack of cooperative purchasing arrangements, primitive marketing organizations, lack of technical assistance, and lack of entrepreneurial skills.<sup>42</sup>

A closely related problem concerns the development and introduction of progressive or appropriate technologies in order to increase the efficiency of production in the F-sector. Progressive technologies reflect the factor endowments of developing countries more realistically than the more capital-intensive technologies of the West and are better suited to the production scale of this sector. The adoption of more appropriate

<sup>41</sup> Harvey S. Perloff, *Alliance for Progress: A Social Innovation in the Making* (Baltimore: Johns Hopkins Press, 1969).

<sup>42</sup> United Nations Economic Commission for Asia and the Far East, *Small Industry Bulletin for Asia and the Far East* (New York: United Nations, 1971); also, United Nations, *Small-Scale Industries in Arab Countries of the Middle East* (New York: United Nations, 1971). For an analysis of the role of entrepreneurship in this sector, see Peter Marris and Anthony Somerset, *African Businessmen: A Study of Entrepreneurship and Development in Kenya* (London: Routledge & Kegan Paul, 1971).

technologies as well as product standardization and quality control would facilitate subcontracting by C-sector industries in the F-sector. Where feasible, government purchasing policies should encourage this practice through legislative requirements.<sup>43</sup>

Policies in support of a general buildup of the F-sector are essential to meeting the urban unemployment problem. In addition, they would reduce the influence of private foreign capital on the national economy without substantially impairing efficiency in production. (Protected C-sector industries operating under import-substitution policies are notoriously inefficient relative to their counterparts in the West.)<sup>44</sup> At the same time, care should be taken not to overprotect the F-sector, so that the benefits of remaining small do not discourage the evolution of capable industries into more sophisticated organizations for larger-scale production.

5. *Regional development: from urban primacy to balanced rural-urban development.* The concepts of economic growth and development—suggestive of an infinite increase in the powers of production—were invented in the course of the industrial revolution in the West. Formalized in the years following World War II, but rooted in the idea of infinite consumer wants (i.e., in a presumption of bottomless human greed), economic growth came inevitably to be associated with industrialization and, by implication, with urbanization as well. Agriculture was thought

<sup>43</sup> *Promotion of Small and Medium-sized Firms in Developing Countries through Collective Actions* (Paris: OECD, 1969); and S. A. Kuz'min, "The Developing Countries: Employment and Capital Investment," *Problems of Economics* 11, nos. 10-11 (February/March 1969): 1-108. For studies related to progressive technology, see Sara Jackson, *Economically Appropriate Technologies for Developing Countries: A Survey* (Washington, D.C.: Overseas Development Council, February 1972), and Marsden, as well as the work of the Intermediate Technology Development Group Ltd., London; and Eugene Staley and Richard Morse, *Modern Small Industry for Developing Countries* (New York: McGraw-Hill Book Co., 1965).

<sup>44</sup> The following examples (taken from Marsden, p. 116) illustrate the varieties of inefficiency frequently found in the large-scale "modern" industrial sector: (1) The large public-sector shoe factory which operated at 20 percent capacity because it had no means of reaching the small private shoe retailers who handled 90 percent of the shoe trade. (2) The hattery plant which could satisfy a month's demand in 5 days. (3) The woolen-textile factory which had a 10 percent material wastage figure (costing precious foreign exchange) because its management did not know how to set and control material usage standards. (4) The \$2 million date-processing plant which had been out of action for 2 years, ever since a blow-out in the cleaning and destoning unit, because there were no service engineers who knew how to repair it. (5) The confectionery plant which was inactive for most of the year because 80 percent of sales were made during the month of a religious festival. (6) The ceramic factory whose quality was poor as it had to use up 2 years' stock of the wrong glaze, imported in error, because the general manager had too much work to attend to all the details satisfactorily. This was attributable to a lack of middle-management and supervisory personnel willing to assume responsibility, combined with a reluctance to delegate authority on his part—both reflections of prevailing social attitudes. (7) The radio assembly factory whose production line broke down repeatedly because of the high rate of absenteeism among key workers.

to be an inferior form of production in the sense that capital invested in industry was assumed to yield higher returns than investments in farming. This notion has only recently been refuted by a careful review of the empirical evidence.<sup>45</sup> Even so, it lingers on and, in most of the developing countries, continues to inform the directions of economic policy.

It is small wonder, then, that low-income agricultural workers migrate to cities. The urban crisis is, in part, caused by superrapid migration to cities but, more fundamentally, it is the result of a systematic neglect of rural economic conditions. People, even in the developing countries, have to eat. Yet the bulk of national investments continues to be placed in cities (and, to a certain extent, in primary production for export), while the volume of imported food stuffs continues to rise.

The aims of a countervailing policy of rural development would be not only to achieve self-sufficiency in food production, but also to expand agricultural employment, to increase the number of rural cadres with special skills in health-related activities—education, building construction, and agricultural production—and to expand employment in middle-sized cities and rural service centers.<sup>46</sup>

Two principles should undergird such a policy: (a) the principle of regional rural development<sup>47</sup> and (b) the principle of tertiary urbanization.<sup>48</sup> The former argues for the coordination of rural development activities on a regional basis (the region in question is generally a small area whose economy is articulated through one or more urban centers), while the latter emphasizes both leading and supporting roles for middle-sized cities serving rural populations. Such roles would include the expansion of services in communication, transportation, health, education, and government administration; cooperative organizations and wholesale distribution; warehousing; agricultural processing facilities; traditional manufacturing (construction materials, leather goods, pottery, metal products); marketing; lodging; and repair services, among others.

A policy based on these principles would help to accomplish six major objectives: reducing rural migration to large cities; utilizing rural labor to the fullest possible extent; increasing domestic food production;

<sup>45</sup> Yujiro Hayami and Vernon W. Ruttan, *Agricultural Development: An International Perspective* (Baltimore: Johns Hopkins Press, 1971), pp. 67–85.

<sup>46</sup> On agriculture's capacity to absorb additional labor force, see specifically Shaw, and the literature cited therein. In general, the conclusion of current studies suggests, with all the necessary caution, that direct employment increases in agriculture require new land settlement and colonization schemes. This option is more feasible in Africa and Latin America than it is in Asia. In any event, the issue is extremely complex and no simple answer will suffice. On our part, we recommend a rural strategy based on small towns and cities and, among other things, we believe that rural development cadres can absorb substantial numbers of youths who otherwise would drift to big cities.

<sup>47</sup> Ranaan Weitz, *From Peasant to Farmer: A Revolutionary Strategy for Development* (New York and London: Columbia University Press, 1971).

<sup>48</sup> E. A. J. Johnson, *The Organization of Space in Developing Countries* (Cambridge, Mass.: Harvard University Press, 1970).

increasing rural incomes (partly through off-season employment in nearby cities); expanding rural markets for urban products; and integrating the national space economy through a tiered hierarchy of dynamic urban centers. Because a large proportion of F-sector employment is already located in middle-sized cities and small towns, such a policy would be further supportive of a strategy that gives priority to the development problems of an urban-linked rural economy.<sup>49</sup>

6. *Population policy: from high rates of natural increase to stabilization.* The employment needs of a population growing at 2.5 percent a year are vastly greater than those of a population that increases at only half this rate. At the former rate, population would double every 24 years, while the latter would extend the doubling rate to about 100 years. Clearly, the rate at which a population grows makes a difference for a nation's capacity to absorb its rising labor force productively.

The very high rates of urban growth that are typical of developing countries do not, surprisingly, lead to rates of urbanization (i.e., to an increase in the proportion of the urban population) that are substantially higher than those which have prevailed historically in the West. Indeed, between 1950 and 1970, high rates of natural increase help to account for the observed increase of 434 million in the world's rural population.<sup>50</sup>

Under conditions such as this, the probability of coming seriously to grips with the problems of urban unemployment and poverty is virtually nil. The fundamental question, however, is the place a population policy should occupy in the entire repertoire of a country's development policies. So long as a policy of stabilizing population is promoted within the context of a government supported and sustained by foreign powers, it will come to be regarded as yet another attempt to tighten the noose of dependency. Within a context of national autonomy, however, a similar policy would assume a critical role in reducing the pressures of population on the country's physical resources. A national objective of maximizing human potentials must penetrate down to the family, where children have traditionally been regarded as providers for old age. This function of children becomes superfluous in an active and mobilized society; instead, the resources of family and community will be used to raise the capacities of young people to their full potential. And in this perspective, high rates of population increase are counterproductive. If migration to cities is to be reduced without affecting the rate of urbanization, the present rate of

<sup>49</sup> "The Philippine census shows that while one-half of the persons engaged in large units are located in metropolitan Manila, this is so only for one-fifth of persons engaged in small units. This means that, proportionally, small firms are located in the smaller cities, towns, and villages, close to the sources of raw materials, employees, and consumers of the final product" (Oshima, pp. 167-68).

<sup>50</sup> Kingsley Davis, *World Urbanization 1950-1970*, vol. 2, *Analysis of Trends, Relationships, and Development*, Population Monographs Series, no. 9 (Berkeley: Institute of International Studies, University of California, 1972), p. 58.



population increase must be cut back through measures that regard the reduction of the birth rate as important an objective as the reduction of mortality.

### Concluding Comments

The six directions of policy outlined in the preceding section are conceived as interlocking elements of a single, coordinated strategy for a country's development in its urban and regional dimensions. The immediate aim of this strategy is to create conditions favorable to full employment of the country's labor force; its ultimate aim is human development in the context of a just society. They imply a radical departure from existing social forms of organization that in most of the Third World are still constrained to the profit motive and accept the principle of inequality as their implicit foundation. Specifically, they are designed to avoid the growing proletarianization of the urban labor force and the persistent pressures to subsistence in the urban economy. If adopted and followed through, they will meet other national requirements as well: increasing savings, minimizing the use of foreign exchange, encouraging spatial integration, and inducing social mobilization for development.

They include the following changes in direction: (1) development objectives: from maximizing growth in GNP to maximizing human potential; (2) economic organization: from a system based in inequality to one of greater equality and social justice; (3) external relations: from foreign dependence to greater national autonomy and self-reliance; (4) industrial organization: from import substitution to an explicit policy of industrial dualism; (5) regional development: from urban primacy to balanced rural-urban development; and (6) population policy: from high rates of natural increase to stabilization.

Policymakers have typically regarded the city as a passive receptacle of development processes that are national in scope; they have treated the city as a social rather than an economic problem; they have responded to the pressures for urban services in cities undergoing the most rapid expansion (usually the principal centers of economic and political power) rather than guiding the urbanization processes in accord with national objectives;<sup>51</sup> they have approached the problem of urban development narrowly from the standpoint of the individual city, unrelated to its region or to other cities; and, above all, they have ignored the 30-70 percent of the urban population—the invisible proletariat—inhabiting the tenements and squatterments of the city.

But the city is more than a passive receptacle; it is tied to the rest of the national economy in many ways, influencing its development as much as being influenced by it. The problem of urban labor absorption therefore

<sup>51</sup> John Friedmann, "The Spatial Organization of Power in the Development of Urban Systems," *Development and Change* 4, no. 3 (1972-73): 12-50.

requires a national strategy and a national approach. It is not likely to be solved by what can only be called tinkering with the urban economy. Both the context of policy and the institutional framework through which it works must be changed. This is the fundamental requirement of what Karl Mannheim calls "substantial rationality." But the basis of substantial rationality is more than a simple aggregate of functional objectives; it is an ideology in which each action acquires a contextual meaning.

At the core of an ideology appropriate to the needs of developing countries lie the ideas of social justice and human potential. As long as these ideas are held firmly in view, many varieties of social ideology are conceivable. Some will argue that the choice we present sacrifices long-term growth and efficiency for immediate gain, that economic development calls for a tough-minded acceptance of a social reality in which it is given for the few to command and for the many to obey. We reject this argument. On the contrary, the relation is one in which the extreme concentration of power implied in the "tough-minded" approach leads to a constriction of freedom, to violence, and to ultimate social disaster. The long-term growth and development of a society, including its economy, is rather the consequence of a commitment to human development and social justice in the present. It is by paying attention to elementary human needs and by abjuring the instrumental use of human beings that the construction of a truly great society can be approached.



---

**Part III Formulating the Issues in Regional Policy**



## Introductory Note

Our understanding of territorial structures and processes is still so limited, and our sense of national purpose for policy so vague, that there is yet much to be learned about the social ills and opportunities that policy is to address as well as much debate about the strategies to be followed. The selections in this part represent an exploration into diagnosis and prescription. They illustrate, we think, that policy formulation is a process of *social learning* where understanding of structure and process interact with diagnosis and rethinking public purposes, and where both interact with the continuing search for solutions.


The first selection, John Ehrlichman's White House memorandum, "National Growth Policy," is an extraordinary document consisting of dozens of questions addressed by the then principal domestic aide to the President of the United States to all departments of the federal government and to scholars in the field (chapter 22). It represents probably the most comprehensive and explicit probing of national policy issues for urban growth and regional development ever undertaken by a high official of government. Wilbur Thompson takes up this challenge (although his piece appeared before the Ehrlichman memorandum) and takes his readers through a grand tour of national urbanization policy for the United States, applying in masterly fashion what is known about the dynamics of urban growth (chapter 23). Tom Herman's brief article (chapter 24) illustrates a little understood political dimension of territorial policy. He reports on the paradoxical opposition by southern federal legislators to proposed legislation that would have greatly aided poor people in their region. It was its feared impact on the existing social structure in the southern states that gave rise to the unexpected hostility that Herman notes. This illustrates many of the conflicts discussed in the introduction to this volume.

The next two pieces deal with growth center concepts and strategies. The idea that public investment programs will have maximal effects on regional growth if concentrated in a small number of favorable locations has dominated discussions on regional and urban growth policy for the last decade. Unfortunately, the discussion has been hampered by a confusing welter of terminology and concepts. D. F. Darwent provides a careful review of the origins and confusions of the growth pole idea and helps clear the ground for subsequent research (chapter 25). In contrast to *poles*, which are nonspatial, growth *centers* are physically located in geographical space. Niles Hansen concentrates on the policy question of how to identify and choose growth

centers (chapter 26). Specifically, Hansen argues the case for the selection of intermediate-size cities as focal points for government policy, but the principal value of the piece is that it illustrates the intuitive basis of debates on this policy issue.

Yet another aspect of regional policy is the choice of region for the administration of government programs. John Friedmann and Barbara Stuckey explore some of the bases and contradictions of regionalization for transportation and other territorial phenomena, the variety of alternative regional definitions and concepts, and the needs of this sector within itself and with other governmental policies and programs (chapter 27).

Part III concludes with two articles by William Alonso. The first (chapter 28), building on the work of Hirschman (chapter 6) and Williamson (chapter 7), focuses on the problems and experience of developing countries and on the dilemmas they face in view of national objectives for economic growth and greater equality. The second (chapter 29) examines in its first part the general purposes of national policy that must be balanced in the formulation of national territorial policy; in its second part, it reviews the territorial dimensions of policies that are not aimed primarily at territorial questions and concludes that the impact of these, together with the forces of the social system, tends to swamp the effects of direct territorial policies such as those directed to depressed areas. It concludes that national regional policy, to be effective, must consider the geographic consequences of all principal public policies. Although Alonso examines the experience of the United States in particular, his conclusions are valid for all countries.



---

## 22 National Growth Policy

John D. Ehrlichman

*Editors' Note:* In 1972, President Richard M. Nixon was considering the possibility of making territorial policy a keynote of his second term in office. The White House staff, of which Mr. Ehrlichman was a principal member, framed the questions below to assist in this decision and asked for answers to them from all major departments of the government and a few scholars. To date, the commitment to national territorial policy remains a continuing but minor theme, and no strong new departures have been undertaken. The questions nevertheless retain extraordinary interest as revealing the manner of thinking of men at the very center of a national government. Any reader of this volume should find it challenging to see whether in his mind he could provide satisfactory answers. Even questions that appear naive will test the sophistication of the response.

One may wonder how a matter of geographic distribution comes to be called "national growth policy." The story is interesting. From 1969 on, Daniel P. Moynihan, President Nixon's urban affairs specialist, counseled about the need for a visible "national urban policy." This was modified to "national urban growth policy" to make it sound more vigorously optimistic. Then it was thought that it would be impolitic to slight rural areas, and the name evolved to "national urban and rural growth policy." This became far too long to say comfortably and the custom developed to speak of a "national growth policy," much to the confusion of the subject.



THE WHITE HOUSE

Washington

May 31, 1972

**Memorandum for Secretary Peterson**

FROM: John D. Ehrlichman

SUBJECT: *National Growth Policy*

There is continuing and growing interest in the subject of national growth policy. In the short run, we can expect issues to arise in the coming months and in the formulation of the FY 74 budget which reflect the Administration's conception of national growth policy. The overall framework is defined in this year's report, and we must continue to define and refine this issue.

In addition, we should begin processes now which will lead to a firm basis for the 1974 growth report.

With these two objectives in mind, I would like your detailed and thoughtful consideration of the specific questions in the attached paper. This document is designed to do the following:

1. To identify areas where agreement among agencies exists and can be fairly said to rest on a firm basis.
2. To identify areas of disagreement among agencies based on different interpretations of trends, facts, studies, etc.
3. To identify important areas where studies or evidence have not yet been developed.

This year's national growth policy study demonstrates a need for a new approach which should be reflected, as appropriate, in agency staffing. Would you please contact my office within ten days with the name of the senior policy officer who will have overall responsibility for this task.

May I please have your response by July 30, 1972. We will review these responses with the President before deciding on further steps.

Attachment



**Scope of Work—National Growth Policy**

In order to develop a more firmly-based position on national growth and national growth policy, the Federal establishment must do much more than it has done in the past to:

Define "national growth policy"

Define the Federal role in national growth policy

Examine present Federal policies toward national growth

Examine present programs with national growth impact

Examine related topics of importance to the process and pattern of growth.

In each of these areas, a number of specific questions must be answered.

**A. Define "National Growth Policy"**

1. a) Is there a consensus in the literature on what "national growth policy" means?
- b) If there is a consensus, what is it and who are the groups and individuals responsible for it?
2. a) How far have the "growth experts" gotten in *specifying* the content of "balanced and orderly growth" in real terms (the definition stressed by the Congress)?
3. a) What evidence suggests that the development of a national growth policy is possible?
- b) Are there examples or studies indicating *in operational terms* what a national growth policy could or could not do?
- c) Can we draw lessons from the experience of other countries?
4. a) Is it possible to limit the definition or concerns of national growth policy in such a way as to make it operationally effective and useful?
- b) Specifically, what kinds of institutions, actions, processes or procedures at the Federal level directly flow from having stated a national growth policy?
- c) What would be the form of stating a national growth policy?
5. a) Would it be *wise* to attempt to develop and utilize a national growth policy?
- b) What are the *political* plusses and minusses of stating such a policy?
- c) To what extent would a national growth policy be subject to economic, political and social constraints which would make the policy irrelevant?
6. a) Are there studies which show the relationship of a national growth policy to other national policies of interest such as economic stability at full employment, industrial planning, tax policies, population growth and movement, environmental and resources planning, etc.?

- b) Who, if anyone or any institution, has, at this point in time, sufficient technical expertise and politically viable organizations to relate these policies?
- c) If no such vehicles for achieving interrelationships exist, what options exist for filling the gap?
- 7. a) Is there any reliable estimate of alternative futures for the country in the absence of a national growth policy?
- b) Can we say with any certainty what the impact of the presence or absence of a policy would be?
- 8. a) Do we know enough about how the general public (not the "growth experts") wants the country to look in the future and, more importantly, what the public is willing to sacrifice in order to achieve it?
- b) If the answer is "yes," what is the source of our information?
- c) If the answer is "no," how do we find out?
- 9. a) What evidence is there to show either a public consensus, need for, or an overriding necessity for the formulation of certain national growth policy objectives?
- b) Are there growth objectives which can safely be said to have widespread or universal support which are not being promoted by present Federal policy?
- c) Do these include the stated Title VII objectives?

**B. Define the Federal Role in National Growth Policy**

- 1. a) What guidelines or concept should serve as the basis for Federal action to achieve and utilize a national growth policy?
- 2. a) How do we define the national interest—how do we balance the relative weights of economic, social and other considerations?
- b) What studies or evidence tend to support various alternatives?
- 3. a) What is the appropriate forum for formulating the Federal role in national growth policy?
- b) What precisely did Congress mean in enacting the Title VII language establishing the national growth policy report and objective?
- 4. a) To what extent should popular present concerns such as population growth and the environment indicate a need for a Federal role?
- b) To what extent should less direct but equally pressing concerns such as trade deficits, industrial productivity and the market economy influence Federal policy toward growth?
- 5. a) To what extent should a national growth policy attempt to achieve welfare and social goals?
- b) Specifically, which goals warrant more Federal promotion?
- c) Is there evidence to suggest that the Federal role must give more weight to welfare considerations than state and local governments give?

6. a) Is there a consensus on whether or not the Federal Government should be concerned with the location of activity in this country?
- b) To what extent should the Federal Government be concerned with the future "map" of the country?
- c) Is it possible to have a growth policy which does not specify where growth should occur or what appropriate actions are needed to achieve locational growth objectives?
- d) How does one distinguish between a "locational" growth policy and any other?
7. a) What are the political pressures at the Federal level which can be expected to bear on the formulation of a national growth policy?
- b) How do these pressures affect Congress and the Executive Branch?
- c) What are the primary interests, both bureaucratic and non-governmental, which will affect the way national growth policies are formulated?
- d) Can these pressures be offset against each other in order to preserve the integrity of whatever policy is adopted?
- e) To what extent can the individual Representative's interest be reconciled with any meaningful national growth policy?
8. a) What does the evidence suggest about the feasibility of coordinating Federal activities to implement a national growth policy?
- b) What has been our experience to date with past or similar efforts?
9. a) To what extent is "disorderly growth" a national growth policy concern?
- b) Is there a national consensus on what constitutes disorder?
- c) If so, what are the specific features?
- d) Are we imposing higher standards on metropolitan development?
- e) What processes of transportation and land development could be changed to accommodate higher standards?
- f) What Federal operational policies does this imply?

#### **C. Examining Present Federal National Growth Policies**

1. a) What evidence exists to indicate how much leverage (potential or real) Federal policies and programs have on the growth process?
2. a) What do we get for the \$35-40 billion now spent annually on Federal grants to states and localities?
- b) Is there evidence to suggest that the leverage which this money possesses has been inadequately used in the past?
- c) How could it be increased?
- d) Is there evidence to suggest that these Federal funds have little or no leverage effect?
3. a) Which Federal directives, if any, expressed in legislation conflict with one another in influencing growth?
- b) To what extent do conflicting directives cancel each other?
4. a) Are there administrative policies and Presidential directives which

- are either internally inconsistent or conflict with Congressional directives, as far as growth is concerned?
- b) What weight must be attached to these various policies?
  - c) To what extent are they self-enforcing and self-implementing, as opposed to being dependent upon forceful advocacy?
  - d) To what extent are these directives location specific?
  - e) Are there studies to show their location impacts and effects?
5. a) Do we have enough information to be able to identify distortions in free (or competitive) market behavior which result from Federal policies?
  - b) Is it possible to quantify the effects of these distortions in any realistic way?
  - c) Are there analytical techniques which can be used in the absence of hard evidence to estimate the efficiency losses of these distortions?
  - d) Are there studies which have used these techniques to estimate these losses?
  - e) Is it possible to estimate the political costs associated with eliminating those policies with a large negative impact?
  6. a) Are there policies of the Federal Government which are not related to programs but more directly related to the economic, social and cultural "environment" which express national growth policies?
  - b) If so, what are they and are there estimates of the impact of these non-programmatic policies on national growth?
  - c) In the absence of quantitative estimates, are there at least estimates of the direction in which these non-programmatic policies impact?
  7. a) To what extent does a consensus exist that there is now a general lack of Federal leadership in developing priorities and policies for national growth?
  - b) Are there operational roles consistent with conclusions of Section B above which the Federal Government could adopt which would achieve the objectives implied in these criticisms?
  - c) If not, are there adequate and appropriate alternative ways of achieving the same objective, such as inducements to state or local action or private action?

#### **D. Examining Present Programs with National Growth Impact**

1. a) Has an analysis been made of the relative impacts of Federal programs on the rate and location of national growth?
- b) What significant assumptions about "national growth policy" and the Federal role are implicit in these measures and in the identification of such Federal program?
2. a) What consensus exists on how the location of jobs should influence housing programs, manpower programs, transportation programs, etc.?

- b) To what extent is population dispersal and concentration taken into account in these programs?
  - c) In what concrete ways do program management decisions reflect these concerns?
  - d) Could a national growth policy be reflected in these program decisions?
  - e) How could program managers, with a much narrower perspective, be influenced to support a much broader objective in making decisions?
  - f) In what form would the directive have to come (i.e., Presidential Executive Order, legislation, informal administrative practice) in order to have any effect?
3. a) Is there an urban fiscal crisis?
- b) How is it manifested?
- c) To what extent does the presence or absence of an urban fiscal crisis determine Federal funding policies?
- d) Is there evidence to suggest that in the absence of a fiscal crisis, revenue sharing, welfare reform, property tax reform, etc., would represent inappropriate uses of tax dollars?
- e) What Federal urban programs, particularly HUD programs, would be superfluous if either the fiscal crisis were shown not to exist or Federal tax transfer policies were instituted?
- f) What Federal urban programs would be particularly effective if the fiscal crisis were shown to exist?
4. a) Is there an analysis to show the impact of the Federal minimum wage on where jobs are located and who holds them?
- b) Is this something that needs to be dealt with in Federal legislation?
- c) What are the possibilities and prospects for various alternative forms of minimum wage legislation that might minimize any adverse impact on growth that is found to exist?
5. a) What does available evidence suggest the effect of urban renewal on central cities has been?
- b) What does the evidence indicate about the relative priority of urban renewal at the local level?
- c) Is the HUD conclusion that the program has failed supported by public or professional consensus?
- d) If so, what policy changes are warranted?
6. a) Is there now a Federal policy or directive on preserving and expanding the stock of housing?
- b) What are the appropriate criteria to be used in defining the Federal role and developing appropriate programs?
- c) Is there at this point a sufficient evidence and analysis to support alternative Federal housing programs?
7. a) What is the appropriate Federal role in evaluating the effects of metropolitan "balkanization"?

- b) Assuming adverse effects can be shown, what should be the Federal role in alleviating them?
- c) Is a rethinking of Federal participation in encouraging metropolitan government indicated?
- d) If so, what studies and demonstrations, if any, should be undertaken?

**E. Examine Related Topics of Importance**

1. a) Is it possible now to estimate quantitatively the fiscal needs of the cities?
  - b) What evidence, for example, supports New York City's estimate that it requires \$50 billion in investment over the next ten years?
  - c) Are there consistent standards of service delivery and amenity which can be used to objectively quantify urban needs?
2. a) What is the impact of urban employee wage demands on the fiscal position of the cities?
  - b) Is there evidence to indicate excessive vulnerability?
  - c) What organizational or procedural steps, particularly Federal ones, could be taken to deal with this problem?
3. a) What is the impact of local property taxes on the location of employment, residential construction, etc?
  - b) To what extent does this question bear on Federal policies?
4. a) Are there indications that the reduction in migration during the last decade is permanent?
  - b) What factors, if present, would restimulate long-range migration, particularly of blacks?
  - c) Are there appropriate Federal policies which could be developed to encourage or inhibit such migration?
  - d) Should such policies be adopted?
5. a) What are the objectives of a growth center strategy as now employed by USDA, DOT and the Appalachian Regional Commission?
  - b) Are these objectives consistent?
  - c) Are the criteria consistent?
  - d) Is there a need for federally-consistent criteria or should this be left to state determination?
  - e) Could the Federal Government set down meaningful criteria?
  - f) If so, how?
6. a) What are the effects of present Federal policies on racial and economic segregation policies?
  - b) Can these be quantified, either in terms of persons affected or economic costs to society?
  - c) What are the effects of racial segregation on Federal programs?
7. a) To what extent should the Federal Government concern itself with the efficiency of local government services?

- b) Is there a Federal interest in insuring that Federal tax transfer payments are wisely spent?
- c) Does this extend to development of new forms of urban technology and services or merely to a "technology transfer" role between cities through advisory and consultative mechanisms?
- 8. a) What are the effects of present court decisions and other policies relating to equality of public services within metropolitan areas and, perhaps, states?
- b) Are there distortions caused by current methods of financing and providing Federal services which are either of Federal origin or concern?
- c) Should the Federal Government attempt to even out the distribution of services (regardless of its responsibility for existing distortions)?
- 9. a) To what extent should patterns of growth within metropolitan areas, especially as they affect the location of jobs, the decay of central cities and increasing segregation, be taken into account in future Federal program development?
- b) Is there sufficient evidence to indicate how the Federal Government can get a handle on these problems?
- c) If not, what studies could be proposed to examine the interrelationships?
- 10. a) Is there a Federal interest in the harmonization and equalization of property tax rates as between classes of taxpayers?
- b) Do, for example, business concerns pay their way in terms of tax revenues?
- c) What costs are imposed on society by classes of taxpayers which are not taken into account in local tax sources and procedures?
- d) Is there evidence to substantiate claims of inequitable income redistribution inherent in such tax policies?



# 23 The National System of Cities as an Object of Public Policy

Wilbur R. Thompson

Many are impatient today with questions such as: how big should cities be, what mixtures of work should they perform, and where should they be located? To some, this emphasis sounds dated and too closely related to the impersonal focus of the earlier city planner who was almost wholly a physical planner. It is people not places that demand our attention, the new breed of urbanists sternly insist. It is poverty, race, crime and the quality of the environment that are at issue.

This charge cannot be answered without first admitting that at many times and in many contexts—in policy contest with local merchants, landlords, land-owners and public officials—vested business, property or job interests of people and places are in sharp conflict. But in more extended and thoughtful discussion among persons with less vested interest in particular places (academics, with little at stake any place?), it should be possible to reaffirm quickly the primacy of people and then to move quickly on to a consideration of the size, function and location of cities as a means to good ends.

We might choose to save some small place because we wish to save some elderly poor person an expensive and unnecessary move, or we may wish to conserve a stock of good, used housing during a housing shortage. Or we may elect to shut down a small place and force migration, to break a cycle of poverty that threatens persons yet unborn. We may limit the size of a city or

empty out a fragile natural area to avoid creating an ecological 'hot spot' that would be very costly or impossible to reverse. We may see in the settlement pattern of this generation the welfare of the next generation. It is not a question of people versus places but rather a matter of the length of the planning period; land planning in a national interest context has a very long pay-off period, greater even than educational investments in most cases. It is in the spirit of the very long run national interest that this paper is written.

## On City Size

More has been written in the past and probably more will be written in the future about city size and social welfare than about almost any other dimension of the city. And the implications of city size commands a large proportion of the content of many papers at this conference. Marginal productivity theory suggests therefore that this effort could find its major focus elsewhere. The direct treatment of city size here will, accordingly, be brief and be concerned with only a very few, select attributes of economic scale that seem to have missed the attention they deserve. The immodest claim is advanced that the points of interest below are more than bits-and-pieces, and may even come to be 'where it's at'.

Reprinted from *Urban Studies*, vol. 9, no. 1 (February 1972), pp. 99–116, with permission of the author and the publisher.

**City size and consumer welfare: variety versus economy**

As a matter of consumer economics, the gains from larger city size probably vary systematically with socio-economic status. Larger local markets bring variety for greater consumer choice, and sheer numbers of sellers promote competition in price and quality. (The small town has always provided the best classroom illustrations of monopoly, aside from the regulated public utilities—'local monopoly'—how many small towns have a 'discount house'?) Lower income households probably gain most from economies of scale in the production and competition in the marketing of standardised goods and services, with increased size of the local market. But the favourable effects of scale and competition on the price and quality of 'necessities' probably flatten out rapidly at moderate city size (100,000 population?). Further, diseconomies of scale may begin to set in at a population of about one million for special socio-economic classes, such as the ghettoised poor. The deterioration of public transit tends to confine the poor to a sub-market of the great city and renders them captive to small neighbourhood stores, the big city counterparts to the local monopolies of the small town, made especially pernicious by the more impersonal and more exploitive nature of insulated 'absentee ownership'.

If we would hypothesise that *economy* in retail trade and personal services is 'U' shaped in relation to city size, we would find it even easier deductively to see *variety* increasing with scale, virtually without limit. Further, in sharp contrast, the advantages of variety would seem to vary directly with education and income; it is the more affluent professionals who have the most to gain from great range of choice. It seems easy, at least to this observer, to arrive at the tentative hypothesis that lower income households, oriented more to economy in standardised goods and services, ride the city size-consumer welfare curve first upward and then downward. Again, the consumer welfare of higher income households would seem, in general, to rise with

increased city size, if not forever, at least much farther.

None of this denies that, with great city size, spillover costs of air pollution and congestion-delay begin to mount and can come to outweigh first economy and ultimately even variety. But the line of argument here is that these increasingly appreciated diseconomies of size come to dominate in the net consumer welfare balance of first the lower and only later the higher income households. In any event, certainly 'economy' and probably also 'variety', as functions of city size and as determinants of consumer welfare by socio-economic class, provide hypotheses every bit as testable as those involving 'external diseconomies' in all its subtlety.

To these more objective variations in consumer welfare with city size, we must, of course, add the more subjective variations in life style that go along with different city size. Even so, the platitude that we need a wide range of city sizes in which to express a wide variety of life styles seems to miss the main public policy issue: 'excessive' concentration of population. The normal market process has given us a substantial range of city sizes. The problem must then be, if indeed there truly is one, that the number of lower income and/or less-educated households that prefer, or at least do not mind, large city living is less than the number of such households that are 'needed' there—industrially and occupationally linked to the more-educated, higher-income households who prefer large cities. Conversely, too few of the high status group are willing to live in smaller places and generate work for those less able and therefore locationally dependent on the former. If so, it would then follow that many less educated, lower income households are 'forced' to live in the large cities that the higher status group prefers. But this is a long and tenuous chain on which to rest national policy so basic as that directed toward 'urban and rural imbalance.'

Why has the market failed to resolve so flagrant a perversion of residential preferences, as this line of reasoning suggests? This bias toward bigness could arise, in part, because those who

supply the relatively scarce professional and technical labour have greater leverage in directing industrial location. But perhaps just as important, the great mass of semi-skilled production workers have implicitly given up whatever influence they might have had on the location of their work (and residence) by quoting a spatially invariant wage through their unions. What is the effective meaning of the Gallup Poll finding that 56% of the respondents would prefer to live in a small town, when they say, in effect, through their supply price of labour, that they do not care where they live?

If it is unlikely that organised labour would be willing to experiment with geographical wage differentials sufficient to induce the relocation of industry, the national public policy issue would seem to be whether alternative ways might be found to register labour cost differentials that would guide production into locational patterns that would better reflect household living preferences (i.e. raise real wages). How much money (after taxes) would various groups of workers trade for the objective and subjective gains from living in smaller places, if a stable way of accomplishing this could be found—cut-throat competition in wages prevented?

#### Urban economic scale and local government

We have had a number of inconclusive studies on the relationship between city size and public expenditures, and by implication public service cost functions and the efficiency of the local public sector. Those local public services that are like, or are in fact, public utilities (e.g. water, sewage disposal and public transportation) do show economies of scale up to populations of medium size (Richardson, 1969, p. 195 f). But the cost functions of the more critical public services (e.g. education and public safety) are much more elusive. Besides, it is more the effectiveness (quality) than the efficiency (cost) of local 'government' that is at issue. What can we say about how 'good' local government is at various city sizes?

Surely, the skill and sophistication of local

public management tends to increase with city size, but then so does the complexity and difficulty of the work. And surely it is not only the lure of higher salaries but also that very challenge of increasing complexity that serves to attract the most able public managers to the bigger cities. But the net balance of the 'supply and demand' for skill is quite unclear. The effectiveness of local government may vary with size due in part to the fact that the strength—bargaining power—of the local public sector, *vis-à-vis* the private business sector, also varies with city size. *A priori*, one might argue that small places are especially vulnerable to the footloose manufacturing firm that threatens flight if, for example, taxes are raised or pollution standards enforced. And, at the other end of the size spectrum, local governments in large, politically-fragmented metropolitan areas feel no freer to pursue tax equity or to guide location and land-use with a firm hand because they fear the loss of tax base through flight as keenly as remote small towns fear job loss. (More often than not, manufacturing firms do not yield a net fiscal gain to small towns because of tax-exemptions or below cost rentals of community owned plants employed to attract them; this is much less true of political subdivisions in large metropolitan areas which ordinarily do tax business property fully and do manage to slough-off some of the costs of servicing plants and worker households on to neighbouring municipalities.)

This all seems to suggest that middle-size urban areas of a couple of hundred thousand population, on up to perhaps a million or so, have the advantage of remaining relatively whole governmentally while becoming moderately stable economically. This city size would seem to provide the strongest and fairest hand in taxation and public regulation. Returning to the prior consideration of attracting high talent into local government, access to 'countervailing power' should count heavily with dedicated and able career public administrators.

One does not need to resolve the relative importance of cost-efficiency versus policy-effectiveness to argue that the latter is more central to a

consideration of national urban policy. A responsible Federal government should care most of all about the ability of local government to reinforce national policy, or at the very least the awareness to not counteract it. We cannot re-make the country into a monolithic system of middle-size cities, and few would so desire, but the Federal government can act to strengthen the hand of local government at both ends of the size spectrum. The rebirth of metropolitan area government in the form of 'two-level local government' is a step in that direction, presumably a lighter and more measured tread (C.E.D., 1969).

#### **Changing criteria in local labour markets**

Larger city size improves local labour markets in important ways, although some very new dimensions of scale may come to dominate. Remote, small places have always borne the burden of monopsony in the local labour market. We still have 'company towns', especially in early-stage processing industries that stay close to the basic raw material, but an ever more pervasive, nation-wide unionism is equalising bargaining power in these places. A monopoly-monopsony stand-off in price power does, however, nothing to reduce the employment instability of a one-industry, one-company, one-plant town which runs the triple risk of a fading product and/or an incompetent management and/or an obsolete facility.

Today, it is more the limited variety and low quality of local employment opportunities in the smaller urban area that creates cause for concern. The narrow range of occupations that emanate from a narrow product base offers little opportunity for upgrading on the job, a still important alternative to formal education. But probably most important of all, especially in days to come, is the placement problem posed by the highly educated husband and wife team—professional or technical labour in joint supply. If females would resign themselves to roles as teachers and nurses, if they insist on becoming educated and professional, joint placement would be easier. There are schools everywhere and a clinic, if not a hospital, almost everywhere.

But reflect on the vocational problem of the psychologist wife of the chemical engineer whose employer selects a small town plant location (near the basic raw material). The probability of both positions appearing in the same urban place is the product of the two probabilities for that size place. But an even greater scale is required to generate enough positions in each occupation so that there is a high probability that both positions will be open simultaneously, or nearly so. Again, while the chemical company may make provision for continuing graduate education in chemical engineering, what are the prospects for post-degree work in psychology in a remote small town?

#### **On the Outlook for Remote Small Places**

##### **The three fates**

In the absence of decisive public policy, the typical remote small town appears to face one of three fates: depopulation, socio-economic deterioration or economic absorption. Towns in very sparsely populated regions will ordinarily experience out-migration to larger places much in excess of the slow inflow from nearby smaller places. While an out-pouring of local residents can be somewhat slowed, a substantial gross out-migration is inevitable, as simple reflections of the search for occupations and life styles that cannot be found at home, and just plain restlessness. With few 'lower' places to draw from and virtually no basis for attracting population from other similar size places, depopulation is often inevitable and inexorable.

A second set of towns in still densely populated rural areas, especially those with high rates of natural increase (e.g., the Deep South), will tend to experience gross inflows that nearly balance the outflows to larger places. The problem faced by slow-growing towns is that, typically: (1) the educational level of the town is higher than that of the countryside, and (2) those who stay on the farm tend to be the more educated and skilled part of the rural population (i.e., the better farmers), and (3) those who move from the town

to the city come more than proportionately from the upper half of the high school graduating class. Thus the town trades the top layers of a superior educational stock for the bottom layers of an inferior one. Towns in this context remain 'alive' by being transformed into low-skill, low-income, isolated ghettos—underdevelopment traps.

Finally, many 'remote' small towns are nearly ripe enough to fall into the expanding commuting range of the nearest metropolitan area; they will be rescued by transportation improvements before they have time to depopulate to a point of no return. Such a town will tend to become a specialised part of a more complex system—typically a dormitory satellite.

#### **Euthanasia**

The appropriate public policy responses to recent developments in remote small places is much too complex to be detailed in so short a paper and too much beyond the present state of knowledge to be discussed definitively in any length paper. Still, the three fates sketched above suggest the broad outlines of three quite different policies and strategies.

No one should still be surprised by the pervasiveness of absolute decline in population; no statistic has been more quoted over the past few years than the fact that one-half of our counties lost population during the 1950-60 period. Nor will the almost inexorable power of this trend be obscured when the 1970 data have been assessed and popularised; two-fifths of our counties have lost population for three successive decades. Certainly, an official depopulation policy and strategy would be political dynamite, but why is there virtually no academic literature on the graceful abandonment of obsolete places, created by those who are not in a politically sensitive position.

In general, the current position of students of regional development is that capital investments in near-permanent infra-structure—interpreted broadly to range from transportation and utility systems to technical institutes and housing—are most appropriate for small *cities* that have been designated as 'growth poles'. Investments in

health and education, especially in younger persons, increase occupational and geographical mobility and are therefore most appropriate for remote small places with bleak futures—places to be emptied out—as well as the more direct and obvious relocation allowances (Hansen, 1970). As general policy, so far so good; but it is an operational strategy that is so needed and so unclear.

A few illustrations will, however, suggest the kinds of questions that we need to address seriously. At what point in the planned depopulation of a settlement is it appropriate for the next larger unit of government to assume responsibility for the financing and/or the provision of a given public service? Just as the county sheriff was responsible for the safety of the rural population before incorporation, so too the sheriff could resume that function with contraction. Should the state reserve the right—practice the policy—of revoking local government charters at some given stage of contraction?

Turning from institutional arrangements to human resources, the order-of-march in migration, by age and education and whatever, certainly affects the stability of the process. Clearly, the normal sequence of migration in which the younger and the more educated tend to leave first is one which threatens a progressive deterioration—a cumulative disequilibrium—and virtually ensures a 'premature' collapse of the local economy, defined here as an unnecessarily accelerated obsolescence of local capital. Could some significant modification of the normal sequence of out-migration be achieved, at reasonable cost? Would it be efficient to pay premium salaries to a few key talents (teachers, vocational counsellors, doctors?) to induce them to stay on longer (or come) in the contracting community, to preside over the orderly amortisation of both sunk capital and of the remaining work-life of immobile people?

While we may want to slow the exodus of certain key talents, we may want to speed the migration of many others. What is the appropriate rate and pattern of contraction under various circumstances? Depopulation might be

speeded appreciably if the transfer of home ownership under favourable terms were made easier. This is especially so in times of housing shortages and credit stringency when a better job elsewhere may be offset with a much poorer house in the new place. A more liberal national housing policy would seem to be a necessary concomitant of a tougher national settlement policy. True, we do not know much about the euthanasia of obsolete settlements, but it seems that we do not even want to talk about it.

#### Small towns as half-way houses

The policy-problem in the case of remote small places in *overpopulated* rural areas is that the locality is beggared in the act of serving the nation as a half-way house between the fields and the factories. A paradox of migration is that everyone could become better off—the rural migrant moving in, the out-migrant to the city and even those who stay and rise in the local hierarchy, due in part to the out-migration of the strong—the nation, as a simple collection of these parts, would be richer because the allocation of labour would be better, but the locality as a *place* could be worse off. Those who move into a community do not have to be less educated or poorer than those they join to impoverish a community, merely of lower socio-economic status than those they replace—the top of the stock that moved out. We do not have to put places ahead of people to recognise that places are environments for people, who may thrive or languish on the basis of rich or poor personal contacts and public sectors.

We are probably fortunate, as a nation, that localities cannot prevent the in-migration that beggars them, but we err in not providing compensatory payments to these staging areas, not just for the sake of equity in income redistribution but even more to provide incentives to localities to undertake the critical functions of education and acculturation less reluctantly and more effectively. The nation should monitor the gross flows of migration at the local level to know better which of its many local agents in human resource development are most strategically located: typically and paradoxically, those communities

most impoverished in inter-regional trades in human capital.

Can we express national development goals in these localities in income terms? Surely, we are not trying to maximise *per capita* income in any give area because that criteria could be perverted into a beggar-thy-neighbour strategy—skimming off the cream. We are not even trying to maximise the income of the current residents of the area because these are open economies with changing populations. The most general goal that could be defended as in both the local and national interest is to create an environment that will produce the greatest learning experience for all who pass through or stay. More operationally, the strategy might be to ignore average income relative to the nation and to concentrate on moving people upward in skill, and outward in direction, if this place has the primary function of first emptying out its hinterland (migration-shed) and then emptying itself out when that job is done. More than cross-section studies of places, we need longitudinal studies of persons to evaluate the performance of places.

But if this place is programmed to be a permanent settlement, then converging on the national average income would seem to be a legitimate test. The corollary to that proposition would seem then to be that we should not accept, as inevitable and permanent, places that are probably never going to be able to attain an average *per capita* income. Or is this position too extreme? Perhaps not, if we normalise *per capita* income for the local industry and occupation mix and correct for the local cost-of-living, including amenities.

#### From small town to satellite

Being rescued from oblivion by a transportation improvement will not be an unmixed blessing for the (formerly) remote small place. That settlement will almost certainly undergo a radical change as it is integrated into a larger urban system as a very specialised part. For example, local trades and services—the local down-town—could become revitalised by the increased outside earnings of the commuting residents. But more

likely that same transportation improvement that speeds workers to plants on the edge of the now less distant metropolitan area also turns local households more to its less distant regional shopping centres, down that same improved highway. A serious loss (inter-area transfer) of local trade and tax base follows. (The parallel with the effect of urban expressways on downtown business in the large central city is noteworthy.)

A second problem, derived from the first, is that members of the local labour force that previously walked or rode buses to work in the old downtown or former local factory district will not be rescued by job opportunities forty miles away and will keenly feel the decline of local trade and service work, not to mention the *coup de grâce* administered to the local bus system by the new outer-directed personality of the town.

Finally, an old and ageing dormitory satellite runs the risk of just wearing out. As the existing stock of dwellings age and filter down to lower income households, this dormitory satellite will become a low-income enclave, unless new houses are injected into the filtering process. But with the gradual disappearance of the local merchant and professional class there may be very few left in town who can afford new houses. A declining residential tax base may therefore be joined to the loss of the non-residential tax base, and a blighted town with a starved public sector is more likely to be avoided than saved by the coming tide of exurbanites and new businesses flowing outward from the sprawling nearby metropolitan area. 'Rescuing' remote small places with highway improvements, as conscious public policy would have to be well handled to avoid simply creating *barrios*, North American style.

#### Populating empty areas

Finally, for the most part, remote small towns will not be saved, they will be replaced. The exhortation that future generations be not deprived of small town living as a possible choice need not be heeded by preserving old mining towns on scarred hillsides or weathered farm villages on bleak plains. We have in fact long been creating

replacement towns in new forms: retirement communities, vacation villages of second homes, artist colonies and professional and scholarly retreats. This swelling movement toward the 'empty areas' has, however, caught us with little policy and not much more thought. The routine (mindless?) replication of traditional (discredited?) suburban forms threatens to repeat past mistakes in land planning—'sprawl' and heavy reliance on the automobile—and seasonal occupancy does not lend itself to the development of competent government, at least in traditional forms (Durden, 1966).

Some predict that the empty areas will come to be populated or re-populated, much more than proportionately, as the principal residence of executives and professionals who because of growing 'telemobility' can have their desks anywhere and prefer to have their homes in natural environments. Populated by consultants, thinkers, writers and artists, these new-style 'rural' areas could come to exhibit higher average incomes than cities—'income inversion' (Berry, 1970). Such an income trend would be reinforced by the continuing elimination of marginal farmers and the accelerating professionalisation of agriculture. Federal rural development policy would then become very different from that now pursued or contemplated. At the very least, a strong land policy would seem to be critical if we are to extend the options now enjoyed by the few into the future under the pressure of the many. Choosing to live in selected places (e.g. Northern New England, Southern Oregon) could be taken as implied consent to live differently: restricted use of automobiles and trailbikes, reduced waste-making, narrower limits on the alteration of natural land forms and so forth. Fragile minority life styles would have to be protected from majority rules.

#### Growth and Migration

##### The rate of local growth is the change in city size

A national policy on the size distribution of cities that does not also incorporate a position on the rate of growth of local populations, that is,

the rate of change in city size, would soon be obsolete. One might, in fact, argue that local well-being is more immediately sensitive to rate of growth than to city size, and that the 'proper' pattern of city sizes might better be derived from the preferred pattern of local growth rates. But, while there have been many studies of actual urban growth patterns, the literature has little to offer on the optimum rate of local growth in population, or even the analyst's own preferred rate of growth.

One does not have to defend the existence of a precise 'optimum' rate of local growth to argue that growth can be too slow or too fast to be easily or well assimilated. A strong case can be made for local growth in employment and population at about the rate of local natural increase, or roughly the rate of national increase. An average rate of employment growth has the advantage of avoiding the chronic unemployment and underemployment that accompanies slow growth in the demand for labour and/or the debilitation of the population stock that follows from 'corrective' out-migration (of the younger and more educated workers). In the other direction, average growth avoids the chronic shortages and congestion that accompanies rapid growth and heavy net in-migration, following from the typical lag of the local public sector in adding to the supply of streets, sewers, school-rooms and trained personnel.

The case for the 'U'-shaped curve is almost always an easy one to make—the pathology of the extremes—but what we need to know is not only where the peak (trough) occurs but also how flat it is around that point, that is, whether moderate deviation from that peak is important or not. On the location of the peak of the growth-welfare curve, a second approximation to this function would bring in population size as a co-determinant. Some have argued that middle-size places of between 200,000 and one million population have the best of both worlds: scale enough to attain moderate economic stability but still small enough to avoid difficult-to-manage complexity. If so, smaller places would be better off being larger, and the sooner, the better.

Smaller places would then attain maximum social welfare by growing a little faster than their natural rate of increase, that is by suffering greater immediate congestion and shortages to achieve greater scale and choice sooner. Urban areas of over one million population should then trade a little adverse net out-migration—loss of youth and talent—for some additional time in which to adjust their much more complex systems to larger numbers of users, and especially their ponderous (even if sophisticated) public sectors. The local growth-welfare curve would then tend to peak a little to the left of the natural rate of increase for metropolitan areas of over one million population and smaller urban places would peak to the right (i.e., at higher than natural rates).

A third approximation of the 'best' local growth rate would distinguish between the pull of rapidly expanding employment opportunities in the local economy and the push of unemployment or underemployment in the hinterland—or better its principal 'migration-shed'. Both the local and broader national interest are served, under demand-pull, by local growth at an average rate and at least the national interest is served by an even higher rate of growth, that is, by heavy net in-migration. (It is surprising how often localities act as if they believe that rapid growth will rebound to their benefit in almost every way, and how little they appreciate that growth is not a universal solvent but that it produces about as many problems as it solves—housing shortages, traffic congestion.) The 'too rapid' growth of midwestern cities in the post-war period was probably in the national interest in light of the great need to depopulate the farm belt, and the fact that this very rapid urban growth was firmly grounded in on the strong demand-pull of a great backlog demand for durable manufacturers also acted to ease the strain. And yet these cities still show signs of the strain of that boom. In sharp contrast, the above average growth in population in Albuquerque shows more signs of being an under-employment-push from the hinterland, and it is not clear whether the local or national interest is being served by its above average



population growth. This is further complicated by the high likelihood that Albuquerque is somewhat too small to be a strong regional developmental pole and, other things equal, should trade some congestion for greater scale, but it would be better to achieve the benefits of a larger population by being pulled rather than pushed there.

In assessing the welfare characteristics of the rate of local growth, we must always take care to distinguish local from national welfare; a given locality can enjoy gains that beggar others. For example, a very rapid rate of local growth could swamp the local cycle, such that a cycle trough is barely if at all lower than the preceding prosperity peak. A local recession would then be expressed as a reduced rate of in-migration, with local employment holding steady (Thompson, 1965). But since total national unemployment is unchanged, one local economy escapes unemployment only by sloughing it off on to others. Similarly, rapid growth draws in the most mobile part of the labour force and gives to that local labour market an unusual power to adapt to rapid contraction at some later time, if need be, with quick and easy out-migration (Morrison, 1970). But again the nation might be better off if the more mobile elements of the population were more diffused throughout the country, minimising the maximum risk of concentrated and/or structural unemployment. Rapid growth also up-dates and local capital stock, making it newer and prettier (to many) and attracting those sensitive to (or guilty about) seeing blight. And again this can be a beggar-thy-neighbour gain; central city mayors have fought New Towns primarily because they provide a too easy escape from the depressing effects of poverty and blight in the older places.

The gain to the nation from very rapid local growth lies almost wholly in the realm of traditional industry or consumer-oriented economics. A sudden large increase in demand for a product or service produced in only a very few places forces a rapid expansion of those places so that the consumer (the nation) is not kept waiting longer than absolutely necessary. When would the total national welfare be advanced by keeping

the nation waiting a little longer in order to build cities a little slower and a lot better?

#### Migration, city size and the market

Can we trust the market and self-interest to adjust the size distribution of cities through migration? On the whole, social scientists have preferred to work with the *process* of migration because it is a clear act, observable and measurable in a familiar behavioural framework, rather than with the *structure* of city sizes, the culmination of a long and complex chain of historical, physical, technological and institutional forces. But migration is much more complicated than it appears on first impression, for there are many less-appreciated economic linkages between persons to add to the many non-economic forces.

Small places do not empty out as promptly or as fully as they would under a pure market model of behaviour for at least two reasons. First, each wave of out-migrants draws more than proportionately from the more educated, talented and ambitious elements of the local population, leaving behind an ever weaker labour pool from which to draw the teachers, counsellors and leaders of all kinds that must meet the challenge of re-working a harder and harder core of the less skilled and/or more immobile workers. Reinforcing this adverse sequence is the conflict of interest between parent and child. Middle-aged parents with poor schooling and few, if any, job skills that will transfer to the newer, larger place are often better off staying on the farm or in the village through the remainder of their working lives, and on into retirement. But their children have no future in agriculture or in the small place and, in fact, face a bleak future anywhere if they remain in the local school system too long. Inter-generational ties and a conflict of interest, with decisions made by unsophisticated parents, is clearly non-optimal from a social point of view.

At the opposite end of the city size spectrum, very large urban areas probably do not slow in growth as much as they should under free market

direction, because migration is characterised by a decided asymmetry. Large cities appeal more to the affluent because they offer much greater consumer choice, especially to the more educated element seeking the more esoteric offerings, and to the young adult with a thirst for adventure and an unlimited faith in his (her) potential to win big in the biggest arenas. But these are also the more mobile persons, so that those who prefer big cities tend to move to them. With time and ageing, many come to favour the environment of smaller places, but the elderly tend not to move easily due to heavy sunk investments in homes, friends and local institutions and due also to the shorter remaining life over which the money and psychic cost of moving must be recaptured.

Note that one can hardly keep from 'choosing' to live in a larger place; one only needs to not move. A person of retirement age could have lived, as an adult in command of his own location, almost five decades in the same city and have witnessed, even at a moderate rate of growth of 20% per decade, a population increase of 150%—from, say, 400,000 to one million. There is an 'age-bias' in migration toward bigness because those who prefer large cities do tend to act on those preferences and those who prefer smaller places tend not to act. Note also that this age-bias tends to reinforce the 'skill-bias' in migration, referred to above, through which professional and technical workers lock the semi-skilled production workers into their locational preferences for larger urban places.

#### On Containing the Size of the Largest Cities

##### Slowing in-migration versus promoting out-migration

Perhaps the reason that we do not seem to be able to come up with an operational strategy for containing the growth of very large cities is that there isn't any. And we should have been addressing ourselves instead to improving their internal organisation. Most of the emphasis in the movement to place an upper limit on the size of cities is on discouraging newcomers, but the

simple arithmetic of the situation establishes that the problem is more that of inducing native residents to leave town. Natural increase accounts for about six-sevenths of the growth of metropolitan areas of over two million population (Taeuber and Miller, 1969). It seems, moreover, most unlikely that we can much slow the natural gravitation of young adults to the biggest places, and it even seems unnatural to try to turn them away. There are, however, some unexplored possibilities for increasing gross flows *out* of big cities.

Local net out-migration has always been a cause for concern because it almost certainly creates a 'youth drain' and tends also to be a 'brain drain'. But perhaps this follows only for small towns; big cities typically attract young adults, on net. Besides, it is the middle-aged and the elderly who are ordinarily the most willing to live in small cities even if they are not the most able to leave big cities, with the middle-aged oriented more toward middle-size places with good jobs and good schools and the elderly looking more to smaller places where living costs are lower and where it is easy to get around. Retirement communities are, in fact, much over-represented among the nation's fastest growing urban places. How much has the growth of these places already slowed the growth of big cities, and how much more of a redistribution of population could be achieved with a more intensive promotion of retirement communities?

#### Transportation technology and natural ecology

But even if we should begin to approach 'zero population growth' as a nation and zero net migration into metropolitan areas, we must still cope with the horizontal spread of cities. Every transportation advance that increases speed of movement, and thereby the distance that can be traversed in the generally accepted one hour travel time limit, extends the effective radius of the metropolitan area and encloses more land area. The smaller metropolitan areas may often be surrounded by sparsely populated hinterlands

and can perhaps be clearly bounded and therefore precisely defined in internal growth rate. These smaller metropolitan areas tend, in fact, to grow by drawing population out of their hinterland. But the largest metropolitan areas have so great a power of attraction and generate such strong 'spread effects' that they tend to be surrounded by densely populated hinterlands. The expanding New York and Chicago urban regions will literally overrun scores of previously independent towns and small cities over the next decade. From independent city to satellite to contiguous suburb in a decade or two.

A 40% increase in the average commuting speed (radius)—a not unreasonable expectation for the year 2000—would double the land area of the local economy. If, moreover, progress in telecommunications reduces the need to appear in person in central places to one-half of the present level, and if the work week has been cut to three days by then, most of us would need to 'commute' to a large centre only once or twice a week. A resident of a suburb or satellite, sixty or eighty miles from downtown Chicago would probably visit there only twice a week, for all purposes, business and recreational. Whether these 'commuters' should be included in the Chicago area growth and size statistics is an academic point; the critical issues lie more in the land planning of these far-flung, but economically integrated, urban regions. The growth rate and population size of our largest metropolitan areas, taken in aggregate, not only cannot be clearly defined, they seem not even to be operational concepts. Witness those tortured statistical artifacts that the census has had to construct: The New York-Northeastern New Jersey Consolidated Metropolitan Area and The Chicago-Northern Indiana Consolidated Metropolitan Area. If transportation progress is inexorable, larger city size seems inevitable.

Intra-urban land-use planning will, of course, come to revolve more and more around serious considerations in natural ecology. Criteria of density and spacing must include but go well beyond amenities of open space, aesthetics of landscape architecture and convenience of move-

ment, into the new world of waste disposal and pollution and the re-cycling of matter. Physical planners have, of course, long been concerned with the loss of open space in and around cities. But too often they over-played the 'calamitous loss of irreplaceable prime agricultural land' which in some not too clear way was held to threaten our very existence. Their anguish in a time of farm surpluses did not transfer well to their critics (Gaffney, 1964). Tax abatement for near-in farms never did elicit much support, nor would it have prevented many rural to urban land conversions. Care would have to be taken that property tax reductions did not simply subsidise and encourage land speculation. The hard case for urban open space has always seemed to rest primarily on recreational use and secondarily on its use to delineate boundaries more sharply to strengthen community identity and build civic responsibility, and virtually not at all on agricultural efficiency.

An ironic turn of fate may bring forth a new variation on the old theme of agricultural land as urban open space. Experimental design is underway in at least one urban area (Muskegon, Michigan) for dumping liquid household effluent on porous soils (sand dunes) over a period long enough to enrich these soils to good agricultural quality. Waste disposal becomes the re-cycling of matter.

Natural ecological imperatives could come to dictate population densities and urban land-use patterns. Given then the speed of transportation and the maximum acceptable commuting time (one hour?), the urban area may come to be shaped 'from the ground-up'. The current direction of causation would then be reversed: city size would not determine population density and land-use patterns but rather land-use patterns and population density, as bounded by current transportation technology, would determine city size. At least in the formal accounting, the population continuum across space could be broken every hour of travel time and summed into (trivial?) population aggregates if one just had to know summary (and nominal) growth rates and city sizes.

**On Industrial Structure****From nation-wide industry location to local industry-mix**

We have a rather considerable literature on the efficient location of industries, in partial equilibrium, but only scattered sentences on a desirable (undesirable) industry-mix for an urban region, especially in a very long time perspective. The distinction between the industry and regional vantage points can be visualised by imagining a series of industry locational maps, one for each of the many industries, overlaid one upon another, through which a vertical shaft is sunk at some given point; the 'core' so removed would then be labelled Cincinnati or Phoenix. Even if each industry were arranged in space in an efficient way, would the many layers yield 'cores' of industries that were desirable, or even viable, at each of the many localities that sum to make up the national economy? Would we not run the risk that some places would have too many jobs for men and too few for women (Wheeling-Steubenville or even places as large as Pittsburgh) and the reverse (Carolina textile towns)? Would we not find high-wage but cyclically unstable durable goods centres in need of more non-durables or services for stability, just as the non-durable towns lack for the heavier, higher-wage durables? And places with sharp winter peaks, unrelated to local climate, posed against the reverse?

There is good reason to doubt that the market will automatically adjust any and all such distortions with compelling price signals and irresistible corrective processes. True, textile mills and garment shops did move into coal and steel areas to tap the large pools of redundant (cheap) female labour, but there are some very clear weaknesses in the connecting chains. The blending of the capital intensive ('heavy') industries, invariably unionised, with the lighter work of the more mobile ('footloose') industries would not seem to happen naturally or automatically through the market forces. Typically, oligopoly product price power in combination with aggressive unionism raises heavy-industry wage rates which

then 'roll-out' into the other occupations in these high-wage centres. There is scant reason to bring the lighter, lower-paying, non-unionised operations into that area, except perhaps to serve that local market if it is unusually large. Similarly, there is little incentive to locate the high-wage work in the low-wage area because the union wage goes right along with the move and the economies of agglomeration and urbanisation of the larger centre often do not follow.

To the extent that these 'heavier' and 'lighter' industries generate complementary labour demands—male and female, skilled and unskilled, heavy and light physical requirements, offsetting cyclical or seasonal patterns—private cost-minimising location that separates them has important social costs in income inequality through unemployment and underemployment. The appropriate policy question would seem to be: Can we plan industrial 'balance' into local economies at substantially smaller scale (200,000 population?) than that at which diversification occurs 'naturally' through free market direction (one million population?).

**Industrial filtering in the system of cities**

The hypothesis has been advanced elsewhere that industries filter down through the national system of cities. Invention, or at least innovation, takes place more than proportionately in the larger urban areas of the more industrially mature regions, and as industries age and their technology matures, skill requirements fall and the industries become free—and competitively forced—to relocate in the lower wage (lesser skill) areas. The higher a city stands on the industrial skill hierarchy, the younger its industries and the more likely it is to fashion an average rate of growth out of a fast-growing industry-mix counter-balanced with declining shares of that work—the innovation of new work and the spinning-off of work that has become routine. The lower an urban area in the skill and wage hierarchy, the older an industry tends to be when it arrives in town and the slower its national growth rate. Intermediate level places tend to fashion a slightly above

average rate of growth out of growing shares of slow-growing industries, but below this size the positive change in share weakens and erodes to zero, leading to slower than average growth (net out-migration) below about 25,000 population and absolute employment decline below about 2,500 population.

This hypothesis has not been subjected to anything more than very preliminary testing (which did support it) and probably can not be rigorously tested without a more disaggregated data than the U.S. Census S.I.C. two-digit ('industry group') employment figures that currently serve as the basic data of 'shift-share' analysis. But the credibility that even casual reflection and observation lend to this hypothesis does seem to suggest that it would not be premature to give some thought to the implications that a national industry filtering process carries for national policy on the distribution of population. Do we know or can we learn to assimilate well, at the national level, such a spatially discrete hierarchy of skill (education), wage level (income) and employment growth (vocational opportunity)? Can this dynamic variation on the more static theme of central place theory be accepted with mild interest as a theoretical or empirical curiosity, as seems to be true of so much of regional economics, or are there important normative and policy ramifications here?

If industries do in fact filter down through the national system of cities, then human resources would tend to filter up in a complementary way. The textile and apparel towns of the Piedmont area have been able to maintain full employment of a rapidly expanding labour force—have been able to absorb the exodus from agriculture in that region—by capturing an ever larger share of these slow-growing industries. But tight local labour markets have not produced an average *per capita* income, as the more ambitious and talented young adults of that region have migrated out (filtered-up) to various larger places. Worse still, this trading of high talent for low skill work has compromised the long run development potential of that region.

This is, moreover, a double-edged sword in

that the larger, more industrially sophisticated urban areas of the North struggle with heavy unemployment that leads to near unemployability and, in large measure, for lack of low-wage unskilled work. New York City needs low skill work as desperately as a textile town needs skilled work. In general, if natural increase tends to produce a population with a random distribution of talent and ambition, and if industrial filtering tends to sector high and low skill work into distant local labour markets, then massive migration is dictated. But if the more talented, motivated and educated are the more mobile, the net flow is biased toward the larger, higher skill places.

#### Large cities as instruments of National economic policy

The urban hierarchy and the industrial filtering process, as presently constituted, could perhaps be assimilated in a way which would serve the national interest if we were to act to increase the mobility of labour, especially the lesser skilled workers, and to arrange inter-regional fiscal transfer of a magnitude that would ensure that local public services were made reasonably uniform from place to place, to assure equality of opportunity. Or we could instead see in the current trend toward large metropolitan areas the resolution of the problem of balancing local labour markets, and reinforce that trend. If nearly the whole national population were contained within a handful of very large metropolitan areas, each of which would perhaps be more a conjuncture of heavily overlapping local labour markets than a single, indivisible commuting space, then virtually the full range of *occupations* would be within easy reach of nearly everyone. Every area would not, of course, produce every product; inter-regional *product* specialisation and trade would continue. And gross flows of migration would go on, as between 'diversified' universities for example. But a nation of multi-million population metropolitan areas would produce balanced local labour markets as a by-product.

Such a distribution of national population would also greatly simplify national economic policy in many ways. Expansionary monetary and fiscal policy is complicated when significant inflation begins to appear in some places (with leading industrial sectors) before full employment is attained in other places (with lagging sectors). Similarly, repressive monetary and fiscal policy can create serious unemployment in, say, durable goods centres while strong inflationary pressures persist in service centres. Variation in local business cycles cannot be fully explained by differing industry-mixes but the industrial diversification that accompanies larger size would certainly make an important contribution to inter-regional symmetry in the response to national economic fluctuations and their treatment. (It is only fair to point out that the inner city ghettos of our largest metropolitan areas would have to be integrated much better into the regional economy of which they are only nominally a part now before we could make a fully convincing case for the employment stabilising virtues of the very large local economy.)

Again, the industrial diversification characteristic of large metropolitan areas tends to generate similar income patterns, and thereby reduces the need for heavy inter-regional fiscal transfers. Federal policy-makers are then freer to concentrate on programmes or incentives that induce a more efficient land-use pattern or that create a more effective organisation of local government, as they are less bound by equity considerations or income goals. It is much easier to deny an average, rather than a low, income area federal funds if they balk at raising pollution standards, consolidating public services or socially integrating populations.

Related to the previous point but worthy of special mention is the tendency for industrial diversification through the concomitant convergence in educational, occupational and income characteristics, to significantly reduce the inter-regional spill-overs that occur through migration. If large metropolitan areas come to trade populations with very similar socio-economic profiles, a major source of unintended inter-regional

redistribution of income and wealth would be averted, such as occurs when small towns give up the top of their high school graduating class and take on responsibility for displaced rural migrants with only a few years of poor schooling. All in all, a nation of big cities greatly simplifies Federal policy, or at least it would if these large *economic cities* were internally competent and effective as local governments. If we do continue moving toward becoming a nation of very large metropolitan areas (economic cities), then the cutting edge of public policy will be less to find an optimum system of cities and more to create an optimum system of governments.

### On Locational Patterns

#### The seeming drift to the coasts

There are two dimensions of the changing spatial pattern of urbanisation in the United States, at the gross grain of the nation as a whole, that stand out over all others: the emptying out of the centre in the drift to the coasts, and the physical formation of greater linear strips of cities—'megalopolis'. The drift to the coasts has been rationalised as a seeking, in part, of easy access to water recreation, a reflection of both a rising *per capita* income (through an income elastic demand) and a change in taste patterns. More recent in nature, the surprising increase in our imports of foreign manufactures, together with the substantial technological progress in water transportation, probably acts to strengthen the position of seaboard cities. Port sites are, of course, especially favoured in the processing of intermediate products where further manufacturing is required.

Some caution should be exercised in a too easy acceptance of the drift to the coast as an independent factor in the American population distribution. About half of the coastline is not experiencing unusual growth; the South Atlantic Coast has, in fact, lost share of population for three decades now and has exhibited absolute depopulation over much of its length. Again, one still does not encounter significant urbanisation north of San Francisco short of the Puget Sound region,

800 miles away. What appears to be a drift to the coasts is in great measure a simple reflection of the growth of large cities located long ago on the coasts—the momentum of prior urbanisation. What is perhaps more defensible is the proposition that a sizeable population (the scale effect) and location on the seaboard or the Great Lakes are sufficient conditions for rapid population growth to large size and substantial growth beyond even from a large base. But apparently not necessary conditions: The combined rate of population growth over the past decade of the six largest interior metropolitan areas (Pittsburgh, St. Louis, Minneapolis, Dallas, Atlanta and Cincinnati) was 15%, slightly above the national average growth rate and well above the 10% rate registered by the one-million-population-and-over metropolitan areas of both the East Coast and Great Lakes regions.

#### Megalopolis revisited

The great urbanised strips that have spread across the country—'megapolises'—are also probably in part real and in part apparent. Gottman (1961) and others have commented at length on the powerful externalities that characterise the Boston to Alexandria strip. But perhaps equally important in its formation is the simple historical circumstance that the nation was colonised along the East Coast and settlements that were more than a day's journey apart at that time—'separated'—are less than an hour apart with today's transportation—'joined'. No one has compared the current population pattern in these urbanised strips with that which would have resulted if each city had grown for two centuries at the normal (national) rate for that size city. How different is the observed settlement pattern from the sheer momentum of urban growth, taken in isolation?

In any event, strips of cities have the special property of combining a relatively high degree of access to a variety of jobs and goods with relatively close open space. This land form produces overlapping local labour markets; more workers live within commuting range of two or more

employment centres. In so far, moreover, as narrowly channelled movement speeds movement through more capital intensive transportation systems, this further increases occupational and consumer choice in space. Because mobility increases with income, in many ways, one gets the suggestion of a partial resolution to the problem posed in the opening pages. Strips of physically distinct middle-size cities, drawn together in time by very rapid (if sometimes expensive) transportation, might happily offer an escape from physical bigness to some and access to economic bigness to others. And with proper land-use control, open space for recreation can be preserved nearby, perpendicular to the strip. By comparison, small, scattered towns gain closer open space but suffer thinner local labour markets, and large circular cities mass jobs and shops but lose access to the countryside. Strips of cities may not be all that bad.

#### Low density urbanisation

A third spatial dimension of American urbanisation and one that seems destined to become a major policy focus is the steady conversion of 'over-populated' agricultural land into 'under-populated' urbanised regions, ranging from the more heavily populated Piedmont Crescent of the Carolinas to the less populated delta of Eastern Arkansas. The Piedmont population, in aggregate, rivals that of the Atlanta city region, but it is still an open question as to whether its loose-knit form can generate as much developmental power as is characteristic of that more classic urban form. The Piedmont has made a relatively successful transition from primary industry into semi-skilled manufacturing, but the next bigger step to science-based manufacturing or to exportable professional services is in doubt. And this second, more demanding step, must probably be made even more quickly under the pressure of the growing shortage of semi-skilled manufacturing work, as successively more sophisticated operations are being spun-off to industrialising nations all over the world. If this sprawling system of small- to medium-size cities



—the largest is only a little over one-half million population—cannot learn to conduct research, marshal information, make good decisions promptly and sell itself in a 'post-industrial age', a second major population readjustment will become necessary, dwarfing the current agricultural depopulation.

There are half a million people in 'Eastern Arkansas', described as a sixty-mile-radius circle drawn around an imaginary centre. If that population mass could function as an urban region, it would be a middle-size metropolitan area, and belong to a class that has exhibited the strongest rate of population growth, better than national average income characteristics and a relative domestic tranquility. The growth of automobile expressways in this and similar low density urbanised areas has brought about a loose integration of what had been a multitude of local labour markets, at least for the more skilled workers who can afford automobiles. The heavy cross-commuting of such an area places it in some uncertain middle ground between its former life as many isolated local economies and its potential future as a single integrated economy. But this population is not at all comparable in socio-economic status to that which normally has been assembled in a metropolitan area of one-half million population. Eastern Arkansas holds today a residual population: a few who have made it in agriculture (large scale, capital intensive farming) and many more who have not, and a trade and service sector that has grown up relatively sheltered from competition.

There are other notable differences between the classic city and an agricultural area that has weakly and tentatively urbanised. Old houses are scattered far and wide and many (most?) are not served by *any* form of public transportation, except the school bus. The normal filtering down of housing that demands no more of a low income family than that it move another mile farther out from the old city centre, along the same old bus route, here requires that the family change towns. Psychologically, this is probably more a migration than a 'move'. A corollary to this is that 'towns' come to be abandoned, like old neigh-

bourhoods in the large city, as part of the normal process of residential filtering—towns like houses may simply wear out.

Many of these old hamlets and villages should not be recycled with new housing for they are old rural service centres that have long lost their economic function, and have not provided local employment for years. Their continued existence can often be justified only to the extent that they relieve the housing shortage of the region, but they perform this service only at the cost of isolating the families forced to inhabit them. This is an 'awkward age' that must be bridged on the way to some new land-use pattern that relates to some new transportation system, such that a full range of housing—houses of all ages and quality—is arrayed along the principal channels of movement. An evolution from dispersed to linear residential patterns does come easily to mind but other options may also come to mind after harder thought. The principal point to be made is that we must learn how to interconnect half-a-million people, spread out over a dozen counties, to achieve viable economic scale or we must assimilate them as migrants.

#### Reprise: population redistribution in a national system of cities

A number of ideas that seem to dominate the discussion of population distribution policy can be brought together usefully in a simple schematic figure and summarised as follows: (Fig. 1).

(a) The depopulation of rural areas should, at least on theoretical grounds, lead to higher returns per worker and higher levels of money income and well-being, as shown by point A. Logically, rural areas should be left with not only fewer farmers but also the best farmers.

(b) If rural out-migrants were to locate in middle-sized urban areas (D), they would tend to increase the money income (productivity) and probably further increase the real income of the inhabitants, as the growing local market permits greater range of choice in goods, services and occupations. If the place of out-migration is in the stage of diminishing returns and the place of



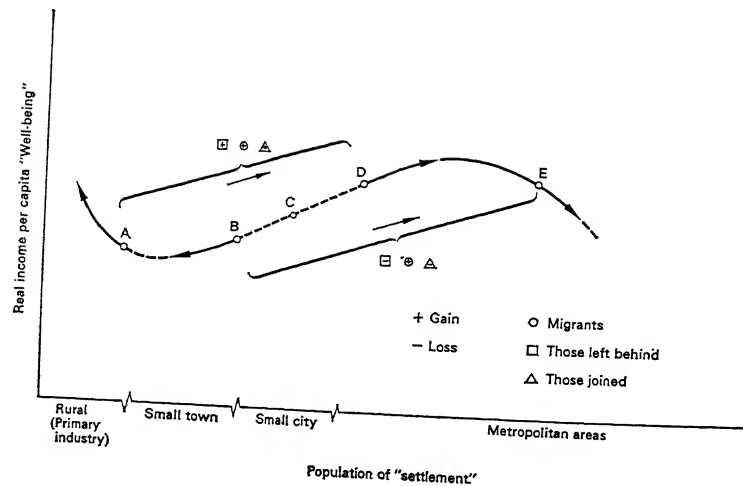


FIG. 1. Schematic representation of the relationship between the 'well-being' of migrants, those they leave behind and those they join, with special reference to the related size of the various populations.

in-migration is in the stage of increasing returns, and wages (income levels) are by implication higher in the latter place, everyone benefits: those who move, those left behind and those being joined. Migration here (from A to D) is clearly in the public interest.

(c) The most discussed case of the day is the migration from small towns (B) to very large cities (E). Such moves usually benefit the migrant who rises from well-being level B to E but could leave everyone else worse off. Those left behind in the small town, growing smaller, face higher costs of utilities and higher taxes for those hard-to-contract (indivisible, heavy-fixed-cost) operations, and a reduced range of choice of goods, services and occupations. The big cities into which the migrants move may have little to gain from greater size but much to lose: increased congestion, greater trip distances, more political fragmentation and inner city housing shortages. If this is the case, then the price signals are misleading; we have made it too cheap to leave small towns and/or too cheap to enter big cities. Migration is profitable for the individual but not in the public interest (Beale, 1970).

(d) The population drift from small towns to big cities has lead some observers to argue for government intervention designed to divert the migrants more toward middle-sized places—small metropolitan areas of 200,000 to one-half million population. These healthy, fast-growing places have demonstrated their attractiveness, vitality and viability, and may also still be in the stage of increasing returns. A larger population will improve their university, museum and theatre offerings and public transportation system. Movements from small towns (B) to middle-sized cities (D) improve the positions of the migrants and those they joined, but do leave those left behind worse-off. And the growing gap between incomes in declining places of 10,000 or less and rapidly growing places of 100,000 or more will induce further migration. A cumulative disequilibrium process is set in motion which seemingly can only end with the abandonment of remote small towns (B).

(e) Perhaps the most popular urban development policy of the day is the 'growth centre' or 'growth pole' strategy. While there are important variations between countries, the domestic

(Appalachian) version most often argued is designed to identify 'small cities' that have growth 'potential' but may or may not develop depending on chance, or may develop unnecessarily slowly and painfully. If, then, migration from rural areas (A) and small towns (B) were coaxed into these marginal places (C), risks and/or delays in their development would be reduced. Such a policy is inferior to the one immediately above in that the former would divert migrants to larger places subject to less risk of failure—investing in smaller *potential* growth centres could result in throwing good money after bad. In addition, the immediate returns to the migrants will be lower if the wage rate in the growth centre of 20,000 population (C) is lower than in the small metropolitan area (D), as is assumed in the curve depicted.

This latter policy is superior to the one above in that, if successful, significant amounts of sunk capital will be saved and social dislocation avoided, as is generally appreciated. What is not explicitly recognised is that the level of well-being (real income) experienced in small metropolitan areas (D) is also a function of their rate of growth in population. And if these middle-sized places are already growing faster than their natural rate of increase in population and must thereby accommodate heavy flows of in-migrants, accelerating that rate of growth may lower real income—housing shortages, schools on half-day sessions and traffic congestion—due to lagging public

investment. Besides, why hurry these favoured places through their golden age?

## REFERENCES

- BEALE, C. L. (1969). *Population Trends*, Part 1. Washington: USGPO, pp. 473-508.
- BERRY, B. J. L. (1970). The Geography of the United States in the Year 2000. *Ekistics*, Vol. 29, No. 174, pp. 339-351.
- C.E.D. (1969). *Reshaping Government in Metropolitan Areas*. New York.
- DURDEN, D. (1966). Use of Empty Areas. *Future Environments of North America*. Garden City, N.Y.: The Natural History Press, pp. 479-493.
- GAFFNEY, M. (1964). Containment Policies for Urban Sprawl. *Approaches to the Study of Urbanization*. Lawrence: University of Kansas Press.
- GOTTMAN, J. (1961). *Megalopolis*. Twentieth Century Fund.
- HANSEN, N. M. (1970). *Rural Poverty and the Urban Crisis*. Bloomington: Indiana University Press.
- MORRISON, P. A. (1970). *Urban Growth, New Cities and the Population Problem*, prepared for the American Association for the Advancement of Science. Chicago: Illinois, p. 24.
- RICHARDSON, H. W. (1969). *Regional Economics*. New York: Praeger. p. 195 ff.
- SUNDQUIST, J. L. (1970). Where Shall They Live? *The Public Interest*, February, pp. 88-100.
- TAEUBER, C. and MILLER, H. (1969). In a statement before the Ad Hoc Sub-Committee on Urban Growth of the Committee on Banking and Currency of the U.S. House of Representatives published in *Population Trends*, Part 1. Washington: USGPO, p. 4.
- THOMPSON, W. R. (1965). *A Preface to Urban Economics*. Baltimore: The Johns Hopkins Press, pp. 176-177.

## 24 Welfare Reform: The Southern View

Tom Herman

ATLANTA—You might think the South would be whooping with glee over President Nixon's Family Assistance Plan.

The \$4.4 billion proposal would pour more Federal dollars into the South than any other region, mainly because the South is the nation's number-one poverty area. Additional money would flow in for years. Some economists confidently predict that the funds would set off one of the region's all-time record shopping sprees, and would also reduce much of the poverty that now discourages badly needed new industry from settling in the area.

So it would seem logical to suppose that Southerners, especially politicians, would be positively delighted at the prospect of all that money.

Logical, perhaps, but wrong. Interviews throughout Dixie suggest that a few Southerners favor the bill, but that the vast majority of politicians and constituents alike agree with Sen. Herman Talmadge, the Georgia Democrat, that the bill "would undermine the best qualities of this nation."

The Family Assistance Plan passed the House earlier this year by a wide margin, but it has run into trouble in the Senate, whose Finance Committee declined to approve the bill. Advocates of welfare reform hope to revive it this week by trying on the Senate floor to attach a modified version to the Social Security bill. But there is no assurance that the measure will pass, even in a watered-down form, and a big part of the reason is heavy Southern opposition.

This opposition becomes more understandable in light of the South's traditional hatred of welfare and its old fear of expanding Federal power at the expense of the states. Many Southerners are sure that the new bill would reward shiftless bums who don't work for a living, a notion vigorously denied by Government officials. But there are other objections that run deeper.

Chief among them is the fear that the bill would speed a reshaping of the economic and social structure in hundreds of Southern communities, especially rural ones, which would receive more aid than anywhere else. And of course a chief beneficiary of that reshaping would be Negroes.

### A Different Approach

The bill's approach to helping the poor differs considerably from present

Reprinted from *The Wall Street Journal*, December 15, 1970, editorial page, with permission of *The Wall Street Journal*.

practice. Most welfare money is currently distributed through the Aid to Families With Dependent Children program, one requirement for participation being the absence of a breadwinner. As President Nixon noted in introducing the Family Assistance Plan, the present method encourages a man with a low-paying job to "desert" his wife and children so they can qualify. The new plan would supplement the income of a low-paid worker, with payments that decreased as his wages improved. The Administration says the incentive to abandonment would be gone.

The new bill also sets a minimum benefit level of \$1,600 a year for the statistically "typical" family of four whose head registers for work and is willing to accept work or training. This may not sound like much of a salary to the average Wall Street Journal reader, but thousands of Southern families earn less.

Sen. Talmadge estimates that the new bill would channel at least some government money to 35% of the population of Mississippi, 25% of Louisiana, and 22% of Georgia. Phil Cawthon, director of Georgia's Department of Family and Children Services, estimates that half of Georgia's counties would discover that 80% of their families would be eligible for some welfare payments. It isn't hard to understand from such figures why Southerners believe this could bring significant changes in the Southern way of life.

A considerable portion of the Southern opposition to the measure is couched in philosophical and pragmatic terms that omit any mention of race. For example, Sen. Talmadge, a leading critic, says, "The Administration has sold this bill to the American people as a work incentive. It isn't. It's a work disincentive. We should pay people to work instead of paying them not to work.

"This nation can't possibly afford it. We'll have a deficit this fiscal year of \$15 billion to \$25 billion. This bill is inflationary." The plan would also, he avers, "make it practically impossible to police the people on welfare to make sure they're not lying and cheating about their status. How are you going to police 25 million people on welfare?"

Similarly, Medford Evans, an official at the Citizens Councils of America in Jackson, Miss., says, "It seems to me that this bill would be a big step toward abolishing the competitive incentive system. If you guarantee a man a certain income no matter how good a job he does, then you aren't giving him any incentive to do good work. The plan is a gradual implementation of socialism and a step toward a slave state where the government supports people and tells them what they can and cannot do."

Herb F. Handley, a Houston real estate investor, says he's against the bill because "it makes it much too easy and too convenient for a person to accept the dole. I worked for every penny I have, and I'm up to my neck with these people who realize they can get more for sitting home and loafing than they can for working. My idea of a good welfare reform bill is one that would make people feel really ashamed to accept a handout."

But of course race is also an important, if not prime, consideration, as it

always is in the South, even when it isn't mentioned explicitly. Many of the Southern objections reflect this racial concern.

One of the bill's most important provisions is the establishment of a Federally funded floor on welfare benefits, which would have the effect of greatly reducing the vast regional differences in welfare benefits under the present system. Some states currently pay as little as \$39 a month to a family of four while others pay as much as \$263 a month. The proposed minimum annual income could have an impact on migration patterns, since many poor blacks now feel obliged to desert the South, where welfare benefits are notoriously low, to live in crowded Northern urban slums, where payments are higher.

Some Southern conservatives fear that this provision might even encourage the reverse—a return to the South by some Southern natives.

"Naturally, it's ridiculous to suggest that the bill will bring home every Southerner who went North and drain the big Northern ghettos, but it does seem possible that some might now decide they can afford to return home," says Paul Mathias, executive director of the South Carolina Council on Human Relations.

Another big worry voiced by Southern opponents of the measure is what all that money would do to the low-cost domestic labor supply. "You're not going to be able to find anyone willing to work as maids or janitors or housekeepers if this bill gets through, that I promise you," says Lester G. Maddox, the peppery Georgia governor.

#### **A Move to Hawaii**

"It's already close to impossible to find people here in Texas to do jobs like housekeeping," echoes Mr. Handley, the real estate investor, "and the bill would only make things worse." Mr. Handley says he is selling his \$325,000 home in Houston to move to Hawaii "just for that reason, because I can't get the proper help for my house here." He adds: "These ethnic groups, and the Negroes, they don't want to say 'Yes sir,' they want to sit down and have coffee with you. This new bill would probably make the situation even worse."

There's also some fear that the Family Assistance Plan would eventually drive up wages. The reasoning is that the vast infusion of welfare money, especially funds given to the "working poor," would bring the annual income of some welfare recipients close to or above the salaries of traditionally low-paid workers in fields like sanitation or possibly even textiles.

The Citizens Councils' Mr. Evans is one who is worried about the effect of higher wages. He says he isn't sure that the measure would attract more industry because "the main attraction of the South for industry in the past has been cheap labor, and this cheap labor pool might dry up if you guarantee a minimum income to all workers."

Perhaps most frightening to Southern politicians, the plan could encourage more blacks to vote and to run for office, by lessening their economic depen-

dence on powerful whites. "There's no question that a fear of economic reprisal has always been a big problem in getting poor black people, and poor white people as well, to vote and to run for public office," says Roger Mills, a civil rights worker in Jackson, Miss.

As Mr. Handley puts it, "You give all these people more money and you'll get more of them voting, which means there'll be more left-wingers running for office and winning because of the bloc vote. And Lord knows, we've got enough left-wingers as it is."

Mr. Evans agrees that the plan "would enormously increase the voting power of the poor people, and in the South an awful lot of poor people are Negroes. They tend to vote in a bloc, and that means there would probably be more Negro officials. It would also tend to escalate the bidding by the national parties for the votes of this bloc."

At present, 67% of the South's 5 million eligible black voters are registered, compared with 83% of the region's 20 million whites, says Marvin Wall, research director of the Voter Education Project Inc., an Atlanta-based organization that funds voter registration drives.

#### **The Impact on School Boards**

The political impact could be greatest in public school boards, since more than 300 Southern public school districts have black majorities. Few of those district boards now have black representatives.

Despite the recent talk of a surge in black candidates for public office, there still are relatively few of them. Mr. Wall says there were about 370 black candidates running for all offices in the 11 states of the Old Confederacy in the fall election. About 110 won, including incumbents, those running without opposition and blacks who defeated other blacks. In the 1968 general election 280 blacks ran and 100 won.

These wide-sweeping implications of the Family Assistance Plan are the main reasons that many people are willing to support it, despite some of its shortcomings. For example, many civil rights groups are unhappy because the support level for a family of four is only \$1,600 a year, well below the \$5,900 set by the Labor Department as adequate.

The present welfare system is far from ideal. In the last 10 years, the cost of Aid to Families With Dependent Children has more than tripled as the caseload has more than doubled. The outlook is no better. A recent study by the Department of Health, Education and Welfare showed that three million people have been added to the welfare rolls in the past eight years, and "at the present rate, at least another four million will be added by 1975."

But the wide agreement over the failure of the current welfare system quickly fades when the talk shifts to how to improve it. The argument isn't over the general notion of helping the poor, since that's become something nobody can afford to be against nowadays. But few seem to agree on who should do the helping, how much help should be offered and exactly who should receive it.

538 TOM HERMAN

It now seems quite possible these disagreements will wind up killing any changes in a system that practically nobody likes.

# 25 Growth Poles and Growth Centers in Regional Planning: A Review

D. F. Darwent

## 1 Introduction

The terms 'growth pole' and 'growth center' have steadily been accumulating a large literature during the past two decades—a literature which is scattered widely across the journals and which attracts contributions from a wide variety of economic and other viewpoints. The terms have, however, been used in vague, indistinct, and distressingly over-simplified ways, there being almost as many meanings ascribed to them as authors writing about them.

It is the purpose of this paper to unravel some of the immense confusion surrounding the notions of the growth pole and growth center, and to evaluate the concepts in terms of their usefulness and their contribution in explanatory and in normative senses. We shall attempt a sifting of sense from jargon, and in a survey of other literature written under a variety of titles show that in some cases the ideas behind the two concepts have been more rigorously developed elsewhere. Initially, we might ask simply in what tense the terms should be used. Are we speaking of places or phenomena that *have grown*, that *are growing*, that *are predicted* to grow, or that (in the normative sense) *we wish to see grow* in the future? Or are we referring to criteria by which growth centers and poles can be distinguished from non-growth centers and poles, past, present or future? It is indicative of the depth of confusion in the literature that there is no immediate answer to these questions. The terms pole and center have been used in all these senses, often interchangeably and often without explicit definition. The resultant confusion is deep and serious.

## 2 Growth poles and economic space

The term 'growth pole' was introduced into economic literature in 1949 by François Perroux (1950), since when it has become associated with an enormous variety of indistinct and ill-defined concepts and notions which have arisen partly from the ambiguity of Perroux's initial formulation, partly from mistranslations from French to English and *vice versa*, and partly from the semantic confusion of many authors. Unfortunately, the concept of the 'growth pole' has a powerful intuitive and emotive appeal which has been exploited by French economists to the neglect of informative, descriptive or analytical content. It has become a magic label and, in the 1950's, a guarantee of publication when used at the head of a paper.

Perroux's initial observation about economic growth, from which much of growth pole literature and confusion has sprung, has often been quoted.

"Le fait grossier mais solide est celui-ci: la croissance n'apparaît pas partout à la fois: elle se manifeste en des points ou des pôles de croissance avec des intensités variables; elle se répond par divers canaux et avec des effets terminaux variables pour l'ensemble de l'économie" (1964a, p.143).

Perroux's meaning in this phrase has caused great confusion. At the outset it must

This paper is based on work done under a grant from the Economic Development Administration of the U.S. Department of Commerce.  
Reprinted from *Environment and Planning*, vol. 1, no. 1 (1969), pp. 5–31, with permission of Pion Limited, London.



be said that he defines growth poles *only* and *specifically* in relation to abstract economic space and *not* in relation to geonomic (or geographic) space, which he dismisses in 1950 as "banal".

Perroux's paper on abstract space defines three types:

- 1 space as defined by a plan;
- 2 space as a field of forces;
- 3 space as a homogenous aggregate.

The growth poles are conceived as existing in relation to the second type of abstract space, and are defined as follows:

"... centers (poles or focii) from which centrifugal forces emanate and to which centripetal forces are attracted. Each center being a center of attraction and repulsion has its proper field which is set in the field of all other centers" (1950, p.27).

Thus, poles are likely to be firms or industries, or groups of firms or industries. At this stage the definition is necessarily broad, in order to encompass a wide variety of the various meanings attached to the term 'growth pole'. Later, the concept will be narrowed<sup>(1)</sup>. It is within these poles that growth and change is initiated, while the connections between the poles, in terms of the flows of inputs and outputs, transmit the forces generated. The poles are therefore best regarded simply as sectors of an economy represented by an input-output matrix in which growth effects can be transmitted across the rows and columns.

Growth in the matrix is directly related to the activity of the poles themselves, and also to the degree of interconnection between them. A condition of 'dominance' of many firms by one firm (or of many industries by one industry) is an important feature of the growth pole notion. 'Dominance' is said to occur when the flow of goods and services from industry J to industry I is a greater proportion of J's output than is the flow from I to J of I's output. In this case, firm I is said to be 'dominant' and firm J 'dependent'.

A further feature of the notion is the emphasis placed on the size of the pole (industry). The rate of growth or change is supposedly directly related to the size of the industry, since the bigger it is, the larger will be its field of dominance over other industries which sell to it or buy from it.

A firm or industry characterized by all three of the above features, that is, high interaction with many other firms, a high degree of dominance, and great size, is said to be 'propulsive' and the firms or industries dominated by it, 'mute' (Perroux, 1964a). This loosely defined concept is a major feature of growth pole notions, and one on which most authors lean. Its ambiguity and lack of clarity are, however, responsible for a variety of interpretations as to the situations which might arise in an economy as a result of action initiated by the propulsive firm. The following description of the mechanics of polarization is therefore a generalized one, based on Perroux's initial formulation, but expanded somewhat to cover some of the ideas of other authors, where these are relevant.

### 2.1 *The mechanics of polarization*

These have been dealt with in theory, though not particularly rigorously, by Perroux (1950 and 1964b), Paelinck (1965 and 1968), Davin (1964 and 1965), Rosenfeld (1962), Boudeville (1966), and others. However, few empirical studies of polarization have ever been made, so the application of theory to reality, in both the explicative and normative modes has been weak. Boudeville's (1957) long paper on the effects of the steel industry on the economy of the province of Minas Gerais (Brazil) is the

<sup>(1)</sup> The problem of distinction between industry and firm can be conceived of as a question of the degree of disaggregation at which analysis is carried out. A 10 digit S.I.C. code would for instance come close to defining individual firms uniquely.

most ambitious. Another contribution is Rosenfeld's (1964) study of the province of Turin.

Perroux (1950) considers the case in which a propulsive firm in a region dominates 10 other firms in the economy. The propulsive firm is given a high degree of dominance over the others, supplying 60% of their inputs and buying 60% of their outputs. A simple matrix is constructed to illustrate this in some detail. Perroux claims that the growth pole notion departs from equilibrium theory by considering global product to be not only the sum of the products of each of the firms in the matrix, but also a function of the effect on a given firm produced by the input and output flows between this firm and all others. He leans very heavily on Scitovsky's concept of 'pecuniary external economies' here. Scitovsky's (1954) paper distinguishes between 'technological external economies' and 'pecuniary external economies'. The first arise when the output of a firm depends not only on its own factor utilization but on the output and factor utilization of other firms. Scitovsky could find only two examples of this in the literature (for instance, a firm benefitting from a labor pool generated by others). The second, 'pecuniary external economies', arise in economic development, when the output of a firm is affected by the *actions* of other firms. The situation in which the pecuniary external economies occur are described as follows:

"Expansion in industry A may also give rise to profits (a) in an industry that produces a factor used in industry A, (b) in an industry whose product is complementary in use to the product of industry A, (c) in an industry whose product is a substitute for a factor used in industry A, (d) in an industry whose product is consumed by persons whose incomes are raised by the expansion of industry A" (Scitovsky, 1954, p.149).

Perroux adds little to this concept of external economies in explaining the mechanics of polarization. He considers the effects of firm A on other firms, but in two specific respects: (a) the effect of A making anticipation of demand, both correct and incorrect, and (b) the effect of A changing the balance of factor inputs in its production function. Thus he shows that an incorrect anticipation of a fall in demand will lead the propulsive firm A to buy less inputs, adversely affecting the smaller firms selling to it; to lay off some employees, causing unemployment, and to raise output prices, affecting firms buying from it. Other effects are also illustrated.

None of this, however, is either precise or rigorous. Moreover it is oversimplified. Perroux assumes that A 'dominates' all other firms, that the regional economy is in a state of full employment of all factors, and that the propulsive firm is very big relative to all others. However, it is not clear what these assumptions add to Scitovsky's concept. Although the assumptions are made, they seem not to be used. They represent differences in degree rather than in kind from those of Scitovsky, in that the effects of firm A on others could occur in a wide variety of economic conditions, and the constraints of dominance, high interconnectivity and large scale merely ensure that the effects will be felt powerfully by the other firms in the economy.

The distinction which it is necessary to preserve in the growth pole notion, between economic space in which poles are defined and geographic space in which they happen to have a location, is a basic and important one which has all too often been neglected. The semantic confusion of attributing to a location the growth characteristics of the pole (industry) which happens to be located there has been made repeatedly. Even when an author does recognize the distinction and define his terms, he often refers to other writers who have defined the term ('pole' in particular) quite differently. The whole of growth pole literature is full of this basic confusion and much of it, in Hansen's (1967, p.723) words, "is badly in need of a thorough semantic reworking".

Paelinck (1965, p.14), in a recent paper, makes the distinction more explicit by considering a region with reference to which poles can be defined. A pole is 'IN' a region when it extends its economic influence over that region (A), and is 'ON' a region when it is physically located somewhere within it (B). Thus, it is perfectly feasible for the industry or firm to be:

- A, B (influencing and located in the region);
- A, B (not influencing but located in the region);
- A,  $\bar{B}$  (influencing but not located in the region);
- A,  $\bar{B}$  (neither influencing nor located in the region).

Moreover, the growth pole concept *a priori* does *not* offer any explanations of the location of a propulsive industry in geographic space, nor of the consequences of a pole having a particular location in geographic space. Aydalot (1965c, p.963) has pointed out that, for example, in considering the firm of Renault (automotive engineering) to be a growth pole, the notion of polarization may help one to understand why Renault is a pole, but it says nothing about Paris, which happens to be its location. The process of polarization, says Hansen (1967a, p.718), "is not amenable to unambiguous geographic location",—a viewpoint with which we must concur.

One of the major points of discussion in growth pole literature, and one which illustrates the depth of confusion in it, is of the effects of the discovery of the gas field in Lacq, S.W.France. French (Aydalot, 1965a) and British (Political and Economic Planning, 1963) economists have either proclaimed its success as a 'pole' without giving sound reasons, or, on the other hand, have claimed that it has failed as a pole because it has failed to affect or induce any other industries to its locality. Yet this discussion is almost entirely misconceived. There is nothing in the original growth pole notion to suggest that the exploitation of gas in Lacq should attract growth *at that location*; this notion claims only that the gas field will induce growth in the economy (without reference to geographic space). It so happens that greater external economies exist (associated with agglomerations) in other locations in France and Europe and, because of this, the gas is transported to those locations. Growth has undoubtedly taken place—but not in S.W.France.

## 2.2 Further developments in polarization

Extensions of the growth pole notion as defined without reference to geographic space have been made by Paelinck (1965), Davin (Davin *et al.*, 1959), Derwa (1957), Perroux (1968), Poittier (1963), and others. Paelinck attempts to generalize growth pole notions into a theory. He follows the concept of Perroux discussed above, and then adds a discussion on polarization in fields other than industrial. Paelinck claims that if the propulsive firm induces growth, either backwards or forwards, this growth can be either a 'lateral' or a 'derived' pole. The example used to illustrate this is the nineteenth century textile industry in Lyons, which induced (backwards) a small chemical industry to provide bleach etc. (a derived pole), and this in turn induced a larger chemical industry providing basic heavy chemicals (a lateral pole). This distinction rests, however, on intuitive rather than theoretical grounds.

Lastly, Paelinck (1965) adds to the process of polarization by analogy, considering other fields in which it might operate. He speaks of "polarization des revenus"—by which he refers to the Keynesian income multiplier effect—"polarization psychologique"—an attempt to consider non-economic, chiefly social, factors—and "polarization géographique"—an attempt to give the whole notion some meaning in geographic space. These attempts are, however, not backed up by theory or empirical verification. They have heuristic value at the most.

Davin, Degeer and Paelinck (1959), writing about the Liège industrial region, and more generally elsewhere (Davin, 1961), use the non-geographic definition of the

growth pole in an attempt to produce advice for the solution of Liège's industrial problems. They identify 'poles' as follows:

"The principal poles are found in heavy capitalized industry and are the domain of large firms; it is essentially a matter of metallurgy involving special types of steel, metal manufacturing industries using the most evolved possible products, chemicals, and activities designed to furnish products for which the demand is in fundamental expansion" (Davin *et al.*, 1959, p.88).

The problem of slow growth (even decline) in Liège is then oversimplified to be a lack of interaction between the major growth poles, which therefore remain potential rather than active. Davin (1961) is not concerned with location:

"la localisation exacte d'une industrie au sein d'une région de développement perd son importance".

His solution to the lack of growth is therefore to stress the importance of exploiting potential linkages between the poles and developing new linkages between them, and with the rest of the economy. He stresses the value of institutional arrangements such as sub-contracting by which the large poles can cause fundamental change in the firms dominated by them. He is also concerned about the enormous influence which the decision makers in large industry have, not only in their own industry but, because of dominance, over the whole economy. Suggestions are made that the state should therefore educate these decision makers in their regional economic responsibilities. To say the least, this is naive and uninformative, and it raises a host of questions, begged by the author<sup>(2)</sup>. In particular, while advocating even greater interdependence between two poles in the economy, Davin forgets the obvious dangers of overspecialization. This is despite the fact that Liège's decline from nineteenth century prominence has, like that of many older industrial regions of Europe, been due to overspecialization in directions which were heavily dependent on world markets outside its control.

### 2.3 Linkage effects

It has been seen above that the process of polarization has many connotations, some of them conflicting, but that a fairly common theme through the literature is that of linkage effects, or interconnectivity, between firms. The study of linkage effects has, however, been pursued much more rigorously in a different field in American literature. Hirschman's (1958) formulation expresses lucidly, at a simplified level, many of the confused ideas prevalent in growth pole notions, though some French authors appear to be unaware of this. Hirschman discusses backward and forward linkage effects in relation to his thesis of 'unbalanced growth' and with reference to underdevelopment problems, but using a completely different terminology. (He does use the word polarization, but in a very different context, with different meaning, which will be discussed later.)

Backward linkages are developed by all non-primary activities, and forward linkages can be developed in all sectors other than that supplying final demand. In trying to develop a system by which to weight particular types of backward linkages emanating from a given firm, Hirschman (1958) speaks of the 'strength' and the 'importance' of the link. The total linkage effect can be measured by the product of these two, where 'importance' is the potential net output of industries which might be induced, and 'strength' is the probability of their coming into existence. This latter can be measured by the quantities of different inputs required by the 'master firm' (exactly analogous to Perroux's 'propulsive industry') multiplied by the minimum economic size of a firm producing each input. Minimum economic size is a concept measured in relation to

<sup>(2)</sup> Yet, curiously, it is this notion of a 'concensus' between government and the major firms in the economy which guides much of French economic planning. Persuasion is more important than coercion.

the local economy taking into account possible foreign competition to supply inputs to the 'master firm'.

In the case of forward linkage, minimum economic size is not a useful concept, "since the size of the markets that might be brought into being through forward linkage does not depend on their suppliers" (Hirschman, 1958, p.102). A guide can be sought in the proportion of the total input to an industry which would come from the 'master firm' (analogous to the concept of 'dominance' in French literature). If this is great, there is a high probability that a small, dependent firm will be induced; if the proportion is small, the probability is low, but the firm will be a larger one (using many other inputs).

In other words, 'strength' and 'importance' are inversely related—where the 'importance' is small the 'strength' will be large, and there will be a high probability of the firm's coming into existence—and this situation characterizes the 'satellite' firm, which is exactly analogous to the 'derived pole' of French authors and has characteristics such as the need for locational proximity to the master industry, the carrying out of only simple transformations and a minimum economic size less than or equal to the 'master'. A good example, given by Hirschman, is the inducement of a multi-walled paper bag producer by a large cement producing unit.

'Non-satellite' industries, in which 'importance' is large and 'strength' weak, are less likely to be induced, since the minimum economic size will be larger and their relationship with the 'master' less dependent. They are analogous to the 'lateral poles'. Thus, it would be improbable that the multi-walled paper bag producer would in its turn induce a pulp and paper mill to supply it with paper. Paelinck's examples, already given, of the nineteenth century textile industry in Lyons are clearly very similar.

This model has considerable heuristic value, as do many of the concepts in growth pole literature. It is, however, far too undeveloped to be of normative value, even in countries of the developing world, and as Hirschman (1958) admits, it is extremely simplified (perhaps over-simplified) in that it deals at most with direct relationships between one firm and a few others. Economic reality is of course much more complex than this, and indirect effects between all firms in a system are of great importance. Moreover, like many growth pole notions this formulation says nothing about location. Chenery and Watanabe (quoted by Hirschman, 1958) extended the idea somewhat by examining interdependency between industries (3 digit S.I.C.) for a number of countries. They were able to rank industries by their degree of interdependence and show that some sectors such as iron and steel have a very high combined linkage, forward and backward, and should presumably induce more industries providing inputs and absorbing outputs than would some other sectors. This approximate ranking Hirschman regards as having some normative value, inasmuch as it adds another criterion for economist-planners to use in relation to developing nations.

However, these simple notions of linkage between sectors across the matrix have been developed in much more detail and with greater specificity by Isard (1960) and Isard and Schooler (1959) as a by-product of 'Industrial complex analysis'. Isard and Schooler pose the problem of a major investment in Puerto Rico. What type of industry would be optimal? The first stage in answering this question is to survey local resources, two of which are the supplies of cheap labor and the proximity of Venezuelan crude oil. These narrow the search to some sort of oil-based industrial complex which can capitalize on cheap labor. However, this still leaves a tremendous number of possible combinations of processes, products and by-products, a field which is narrowed further by looking at the available market for oil-based products and assessing Puerto Rico's comparative advantage. This results in the choice of a

complex whose major outputs would be Dacron (for a wide market) and fertilizer (as an import substitute). The choice made, however, is constrained by the need to take advantage of two major factors not previously discussed, namely:

- 1 economies of scale for each element in the production process;
- 2 localization economies that might result from agglomerating all or some of the processes at one point—offset by transport cost considerations.

By treating these two factors simultaneously with their cost analysis of 'in what' to invest, Isard and Schooler (1959) are also considering location, and an optimum is a theoretical possibility. Intra-industry linkages are therefore considered explicitly in the analysis, this being achieved by the construction of a precisely quantified inter-activity matrix of inputs and outputs. Moreover, by dealing with the problem at a highly disaggregated scale, the authors approach economic reality much more closely than any of the others reviewed in this paper. They are able to deal with links which are not only direct but, by repeated iteration, those that are much more indirect. The 'growth pole' is therefore born fully fledged, both at a location (or locations) compatible with its functions, and with its appropriate assemblage of satellites and non-satellites. The only major element not covered by this technique appears to be that of urbanization economies which might result from the agglomeration of firms in different industries. (The attraction of existing agglomerations on the proposed complex is covered in the analysis only to the extent that sources of labor and absorbers of output might be located in them.)

We must now move on from this review of the non-spatial concept of the growth pole to ask how authors have looked at the problem of the spatial incidence of growth and the spatial allocation of investments within or between regions of a nation.

### 3 Growth poles (centers) in geographic space

#### 3.1 *Introduction*

Growth pole notions discussed so far are independent of a spatial context, as are Scitovsky's external economies and Hirschman's linkage effects. The interaction between a propulsive firm or industry and others is seen only in relation to the matrix of a theoretically open economy whose bounds are arbitrarily limited to a nation or a region. Locations in geographic space are not considered.

However, since all units must have a location, and since in regional economic development the question of 'where?' looms large, then despite the fact that poles are independent of geographic space, their existence within it poses complex problems unexplained by growth pole 'theory'. This situation has obviously concerned French authors, and the result has been the emergence of another group of concepts (also referred to in the literature as 'poles') which we will discuss under the name of 'growth centers', defined as locations (usually cities) in geographic space. Propositions and theoretical descriptive studies of growth 'centers', so defined, have multiplied in recent years and have a strong normative element which has in part been harnessed into the provisions of the fifth French "Plan de Développement" (Political and Economic Planning, 1963).

Yet in the whole of French growth pole literature, only one example (Boudeville, 1957) is known to the author of an attempt to make explicit the connection between the conditions for the existence of a 'pole' defined in abstract space, and the conditions for its appearance and location in geographic space, as a 'center'. Even in the American literature, papers on this aspect are not numerous—only Isard and Schooler (1959) treat it adequately. This extraordinary gap can only be explained by the great semantic confusion of many authors between the industry (pole) and its location (center). This confusion has been so profound that explanations of how a link occurs between

the two, have been thought unnecessary, or at least are presumed to be 'understood' by the reader.

Clearly, as soon as we begin to discuss the location of growth poles in geographic space and in particular the conditions for and manner of their growth, we are dealing with a somewhat different set of variables which intersects only marginally the set discussed in the French literature under 'poles'. The location of the propulsive firm is presumably best dealt with in terms of location theory, as are the locations of the dependent firms induced by its growth. However, it would be unreal to expect classical location theory to say much about this without also considering the amorphous field of external economies in both public and private sectors, and the nascent 'theory of agglomeration' (U.S. Dept. of Agriculture, 1966) in which there are advantages to location and growth in a large city over and above the market pull of an agglomerated population which would in any case be considered in the context of location theory.

### 3.2 *From growth poles to growth centers*

A straight interpretation of pole notions into geographic space would produce naive answers. Firstly, one cannot talk about the location of a propulsive *industry* without assuming that all firms or processes in that industry are agglomerated in one location, which is clearly unrealistic. The optimum location of the propulsive *firm* would (from the firm's point of view) be seen in terms of the location of its sources of inputs and markets for outputs. A simple version of location theory could be envisaged in which optimum location would be a product of the relative transport costs of the different inputs and outputs, weighted by such factors as the firm's labor demands, and the relative closeness of the manufacturing process to the final demand sector. The location of the induced, or dependent, firms would depend on whether forward or backward linkage, and satellite or non-satellite firms, were being considered. In backward linkage, the satellite (defined as being dependent on the propulsive firm as a market for more than 50% of its output) will, depending on the relative weight loss or gain in its manufacturing process, have an optimum location heavily influenced by that of the propulsive firm. Backward linked non-satellites will be less likely to be influenced by the location of the propulsive firm. In the case of forward linkages, similar relationships might be envisaged, with the non-satellites closer to final demand and more oriented towards the market.

This is obviously a very crude statement of locational criteria. However, it illustrates that even at this level the notion of growth pole has very little intrinsic value where location is being considered, and that serious locational considerations would have to be dealt with in terms of location theory, with growth pole notions playing a very small part. In fact, we must turn to Isard and Schooler (1959) for a demonstration of the link between the pole and its location. Again as a by-product of their industrial complex analysis, Isard and Schooler throw valuable light on this issue. By developing the inter-activity input-output matrix for the elements in an industrial complex, they are able to deal with the choice of what to invest in (growth pole), simultaneously with considerations of the optimum size of each of the elements in the complex, the scale economies involved, and the transport costs of each of the inputs and outputs. The linear programming solution to this will give the optimum location for the optimum combination of linked processes in a complex, in terms of the maximization of some objective function—say profits. Scale economies and localization economies are therefore adequately treated. In the case of Puerto Rico, Isard and Schooler were able to produce a solution which also indicated which processes would be best agglomerated to achieve localization economies, and which dispersed (near inputs, or markets) to minimize transportation costs. The analysis, however, as pointed out above, stops short of the consideration of urbanization economies which result from the agglomeration of different industries.

Boudeville (1957) attempted an empirical demonstration of growth pole notions, and of the location of a growth pole in geographic space, by measuring the impact of the steel smelting industry on the economy of the province of Minas Gerais, Brazil. This measurement of 'forward linkage effect' is based on relating the increase in steel fabrication output to a 1% increase through time in steel output in Minas Gerais, and of relating output in the two industries for all Brazilian states at one moment in time. The results indicate that while there appears to be a temporal relationship between steel smelting and steel fabrication, the spatial distribution of the latter is more closely correlated with the size of the population of the largest agglomeration in each state. In effect, while Minas Gerais had one of the largest and fastest growing steel smelter outputs in the nation, its output of steel fabrication products was small. This was correlated with the small size of the state capital, Bello Horizonte (300 000) compared to San Paulo (2 000 000) and Rio de Janeiro (2 300 000), which had the largest steel fabrication industries.

This attempt to separate the 'agglomeration effect' from the 'polarization effect' is thus not very successful, since it is not possible to induce causal connections from correlations and regressions. In particular, even if we accept Boudeville's multiplier effects as evidence of a causal relationship, it is difficult to see how the polarization effect is separately identified. The existence of a steel fabricating industry is obviously dependent on two major criteria—the existence of a market and a supply of steel. Thus it may well be, since steel fabrication was more closely related to the former than to the latter, that a steel fabrication industry would arise even if steel were not produced in Brazil, but imported. Boudeville's consideration of backward linkages are equally inconclusive in this respect.

There are few empirical papers in the literature and since they do not add much to the above we must conclude that not only does the notion of 'pole' have limited theoretical value in a locational context, but that empirical demonstration of the 'inducement' effects hypothesized by Hirschman has not as yet been satisfactory.

### 3.3 *Growth centers and the polarized region*

Despite the flimsy theoretical background and the lack of empirical verification, there is nonetheless a great deal of intuitive appeal in the notion of a growth center in which economic and social development is initiated and transmitted to an area around it. Moreover, the most important normative questions of regional economic development, those concerned with the regional allocation of investments in both time and space, can be given some clearer direction if this intuitive notion is adopted. Thus, for example, such a notion would imply that investment is best concentrated in growth centers rather than scattered around in some vague quest for 'balance' or 'equity'. It might also imply that the existing central place structure of a nation could somehow be adapted to serve specific goals of growth initiation and transmission.

Because of the above consideration, the notion of the growth center has been of interest to many economists. This has been so especially in France, where the system of national and now regional economic planning calls for definition of regions, and advice on the spatial as well as sectoral distribution of investments. This interest in the normative has also broadened the basis of growth center notions to include variables other than economic—such as social, political, cultural, and psychological.

Much of the work in this general direction centers around Boudeville's (1961c and 1968b) conception of three types of region—homogeneous, polarised, and planning or programming, regions. The homogeneous region, beloved by geographers, has maximum internal homogeneity and maximum external heterogeneity in whatever factor is being measured. Polarized regions are defined to be that collection of geographic spaces in which connections and flows of, for instance, goods, services and political



allegiance are predominantly in one direction—towards a central point, or 'pole', which dominates the region. The boundary of a polarized region is therefore that line beyond which flows and connections are predominantly in some other direction towards some other pole. This concept is very similar to Derwent Whittlesley's 'nodal region' (Centre des Études et Recherches Economiques et Sociales, 1959) and has many features in common with John Friedmann's 'center-periphery' model. The polarized region can exist at any scale, and smaller ones, polarized around smaller centers, will tend to 'nest' within larger ones. The idea of the polarized region is therefore compatible with the central-place structure of a hierarchy of cities of ascending size and function, with the 'growth centers' normally being the larger city or cities in the region, at whatever scale is being discussed.

This concept has been incorporated into French regional planning, which has three homogeneous regions—Paris, West and East—nine polarized regions each with a center called a 'metropole d'équilibre' (excluding Paris), and 21 programming regions for short term allocation of resources. The eight 'metropoles' outside Paris have been chosen on the basis of their size (some consist of several cities, such as Lille—Roubaix—Tourcoing), with the aim of decentralizing some of the regional functions from Paris. The goal is to solve the problem of 'Paris et le désert français' which arises because Paris taps off most of the regional functions of the other regions and their centers, so that the small provincial towns are only feebly dependent on the regional centers, and Paris's influence spreads through the nation. The policy is to concentrate investment in the eight regional metropoles, each of a minimum target size of one million, and develop an appropriate 'armature urbaine', that is, a central place structure, which will maximize the interdependency of the peripheries of each polarized region with its respective center by gathering flows from the periphery and directing them up through the hierarchy, and by bringing the benefits of urban life down to the smaller centers. The smallest villages and hamlets in the most depressed areas will be encouraged to decline through out-migration, thus cutting off the 'tail' of the hierarchy. In Friedmann's terms, this is a policy of encouraging spatial integration between core and periphery, at the *intra*-regional scale—the level of the polarized region—while correcting the imbalance evident between Paris (as a core) and the rest of France (as a periphery) at the *inter*-regional scale.

The 'growth center' is, in this normative sense, a somewhat wider concept than we have previously seen, and is not closely related to Perroux's initial 'growth pole' notion. In this sense the growth center has been extended somewhat by other French authors.

Lebret (1961) attempts to link agglomerations with growth poles. He sees agglomerations as composed of one or more 'unité motrice'—assemblies of capital or resources, which by their size and growth can be termed growth poles. Most large French agglomerations have several poles located in them and are able to generate growth by their size, which releases valuable external economies. Growth, however, is considered not merely in terms of the market, but also in terms of the functions and services which the agglomeration can provide. A strategy of development for the economic space dominated by the center is then based on the organization of the 'unités motrices' in such a way that they lead to maximum development at minimum cost, and this will be best achieved if they are complementary and generate a large number of interlinkages and multiplier effects. Lebret considers that this policy is viable at any geographic scale, but principally at the level of the metropoles.

Antoine *et al.* (1968) see the metropoles as being propulsive on account of their well developed tertiary sector, since this is the fastest growing part of the economy. They advocate a policy of investment in the service sector located in each of the metropoles, and argue that this will help, not hinder, the smaller towns of the regions, by generating multiplier effects. The aim is to build up the functions and

level of services of the metropolises so that the regions become oriented fully towards them. Hautreux (1966) agrees with the above view, but stresses that since the central government rarely has much control over the location of commercial services, policy should be concentrated on building up public sector services such as education and administration, which, by increasing the flow of customers to the agglomeration, will encourage in time the development of commercial services. This is also consistent with the French policy of decentralizing public sector services from Paris. Labasse (1968) also stresses the tertiary sector as a basis for growth, in that a highly specialized agglomeration with many services is likely to attract and retain the elites necessary for the decision making process in economic development. The presence of 'rare services' can thus be looked upon as both a cause and an effect of polarization and the integration of economic space around the growth center.

This line of thought has been extended by considering the interaction between polarized regions, in terms of their growth centers, and the poles which form these centers. Perroux (1961) attempts to demonstrate the interaction between Region A with a propulsive firm and Region B without one, but he does this without reference to geographic space, so his speculations are of limited value. Boudeville (1969c), on the other hand, has looked at the relationships between the eight French provincial metropolises in theoretical and empirical terms, and devised a matrix of their bilateral interaction measured by the total number of employees controlled by one city's decision makers in the other seven cities. This produces a picture of relative dominance in which Lyons emerges at the top. (Paris, which dominates all the other cities, is excluded from the matrix.) Boudeville speculates on the possibilities of extending this approach to the social, institutional and political fields by means of various surrogate measurements based on data collected in surveys. By this means it would be possible to identify the groups which, by their power and influence, have some control of the polarization process. Friedmann (1967) has amplified this in a recent paper based on 'center-periphery', discussed below.

In his book on regional planning, Boudeville (1966, p.108) deals with regional models, one of which is based on the polarized region. He develops a model which makes possible the definition of boundaries of regions polarized around major cities by application of a modified gravity model of the information-flow type whose inputs are population and distance only. Modifications are made for particular products and functions based on distance elasticities. An application of the model considers three polarized regions of the Rio Grande de Sul, Brazil. A proposal to relax customs dues implies that the boundaries of the three polarized regions will change. Boudeville predicts the redistribution of population needed to retain the boundaries in their original position, and the required number of extra jobs implied. This is then related to a given amount of investment in manufacturing, and the 'optimum' allocation between the three regions is indicated. This model has very limited applications, however, as the somewhat esoteric example might indicate.

Hirschman (1958) and Myrdal (1957) have also contributed to this aspect of growth center literature. Independently, they both talk of a process whereby one region (called 'North' by Hirschman) is the growth center, being advanced and developed, which influences or controls the rest of the nation ('South') by two processes—'polarization' and 'trickling down'. Myrdal's terms, exactly analogous to these, are 'backwash' and 'spread' respectively. 'Polarization' effects exercised by the North on the South tend to be to the South's disadvantage, and are due to the North's stronger economic position. They include severe competition from the South's relatively inefficient industries, and a tendency for selective migration of the young, skilled, educated people from South to North in search of the greater opportunities and apparently higher salaries available in the latter. Because the North's industry is

productive, what little capital South possesses is also likely to migrate to the North, where interest rates are high and security guaranteed. The favorable 'trickling down' effects from North to South are the increase of Northern purchases and investments in the South and the absorption by the North of some of the South's underemployed, thereby raising per capita incomes in the South. Hirschman shows, however, that these effects are likely to be balanced in favor of the North (the 'center') and against the South ('periphery'), and that this situation of imbalance will tend to continue up to the point where the lagging of the South begins to affect the North's growth, or where the South will seek to redeem the balance by political action or revolution. Hirschman's use of the term 'polarization' is thus rather different from that of French authors, but he is clearly expanding the notion of the growth center into a more broadly based concept of regional economic development compatible with his notion of 'unbalanced' growth. This has been taken up by Friedmann and Hansen.

A somewhat different direction has been taken by Fox (1966) and his associates. Fox defines and discusses growth centers in a very restricted sense. His definition is a normative one, "an urban place which can act as a focal point for development planning", and is related only to the development regions and districts as defined by the U.S. government's Economic Development Act. In this context, "a growth center is typically an urban place of less than 250 000 population which acts as the vital heart of its development district". Fox goes on to discuss criteria by which a distinction might be made between urban areas which are growth centers, and those which are not. Yet even in the very restricted sense which Fox uses, the criteria fall far short of enabling such a distinction to be made. The criteria mentioned by Fox include strong linkage to the national economy, the center of a labor market, a major retail trade area, high level tertiary functions, a large volume of wholesale trade, and good communications, none of which is specific enough to help us separate a growth center from a non-growth center—they are characteristic of most urban areas.

Further discussion by Fox concentrates on identifying the size and rate of growth of a center with the population density of its hinterland. A clear, positive relationship is hypothesized similar in many respects to Chaineau's (1965) model. Normative prescriptions for the induction of growth include various methods of reviewing regional productivity and for reducing the regional capital output ratio by investment in the growth center, and the establishment of policies which will lead to the reduction of differences in the marginal rate of return of factors, the assuring of economically efficient locations for every enterprise, and the concentration of public investment in areas of maximum growth potential. The potential incompatibilities involved here seem not to concern Fox.

#### 3.4 *Center-periphery*

The need to answer the question 'where?' in explanations of the occurrence of economic growth and in problems of regional allocation of investment has been expanded by the work of John Friedmann (1966, 1967a, 1968). In developing the center-periphery model, and in increasing its specificity and level of detail to cover variations within the periphery, Friedmann stays close to the real problems of regional development, the normative questions. His work has very little relationship to the rather confused, ill-defined, vague and often barren discussions of growth poles in French literature viewed above. Indeed, although Friedmann claims the opposite, it is difficult to see in what respects his work owes anything but the name 'growth pole', and its intuitive connotations, to the French authors preceding him. The distance separating Friedmann from other authors is so great that in a recent review of the growth pole concept (Hansen, 1965), his work was not mentioned.

Friedmann's major contribution to growth pole and growth center concepts is to show how deficient these concepts are in most respects. He does this by developing a richer and more satisfying descriptive and normative model of the spatial incidence of regional economic development. Specifically, Friedmann's work is outside the problem of the inter-sectoral distribution of resources, which has been shown above to be the main source from which growth pole notions have sprung. He addresses himself squarely to the real problems of regional development in geographic space, and via his 'center-periphery' formulation emerges with a nascent theory of 'polarized development' which will ultimately cover not only the narrow range of economic variables, but also explicitly social, political and cultural developments in geographic space.

In his book on Venezuela, Friedmann (1966) begins by describing the 'regional policy problem' as an issue of applying national policy in a spatial dimension, and indicates that it emerges as an issue particularly in the second of the four phases of national development:

- 1 pre-industrial;      2 transitional;
- 3 industrial;          4 post-industrial.

He is concerned with the integration of what have, in the pre-industrial phase, been independent regional economies of a colonial nature, usually on a coast and tied economically more closely to their master-country abroad than to each other. Regional policy becomes a necessity when such a nation gains independence and is faced with the problem of integrating these regional economies. It is here that Friedmann introduces his 'center-periphery' model. The independent economies usually have a single center on the coast, and a periphery, loosely tied to this center, from which produce is extracted for export to the master-country (colonial) abroad. Relationships between the center and the periphery will be minimal, and those that are developed will tend to be one-sided, supporting the center at the expense of the periphery, which remains a backward, exploited area unable to grow because it is feeding the growth of the center. This 'dual' notion of the economy becomes imprinted on the national economic space.

Limited evidence is available elsewhere that a center-periphery structure can exist at different geographic and time scales, and at different phases of development. An attempt has been made to view the Western Europe-Eastern Seaboard U.S.A. area as a world 'growth center' until recently living off the periphery, the rest of the world (Chisholm, 1964). Modern France, with its extreme centralization on Paris as a 'center', to the great disadvantage of the 'periphery', can also be thought of in these terms.

Friedmann discusses the interaction of the core and the periphery under eight propositions concerning the incidence of economic growth with spatial implications. These propositions amount to the view that regional economies are essentially 'open' and because of this economic growth is usually induced externally, initial impetus being the export of a primary product or resource to other areas or nations. Successful translation of this export-sector growth into 'residential' growth (internal to the region, and serving local markets) depends on the socio-political structure and the distribution of incomes and expenditures in the region. Within a given region, the growth of residential activities will be enhanced if local investment and government infrastructure are encouraged. However, all this is dependent on local leadership which, unfortunately, is also a product of the region's development experience. This will be encouraged by a decentralized administration providing opportunities for local decision makers.

In spatial terms, economic growth as seen in these propositions would occur in a matrix of urban regions which are the building blocks around which economic space

is constructed, and evolve in the direction of ever greater integration. In effect Friedmann generalizes his 'location points' (locations in economic space) into cities and towns which, because of their valuable urbanization and localization economies, tend to be favored as points of growth. There emerges a hierarchical system of cities, and thus also of urban fields, which is held to be evidence of increasing spatial integration, and therefore development.

Economic growth is related to this system in specific ways such that the population of the sphere of influence of a city will be proportional to the size of that city, economic growth will be a function of distance from the central city, growth potential between two cities will be a product of their size divided by their distance apart, and economic change will tend to be transmitted from higher to lower orders in the hierarchy. Friedmann implies that economic development is closely related to the emergence of a highly developed and interconnected functional hierarchy of cities of the Christaller type, and that growth is in some way proportional to the size of an agglomeration (modified by imperfect labor mobility). In terms of encouraging overall development he therefore advocates such a hierarchy of cities as a means of integrating the periphery with the center, or core. This is in many ways similar to the French notion of the 'armature urbaine' discussed above, and to Lloyd Rodwin's (1961) 'concentrated decentralization'. However, Friedmann goes further than this.

Along with Perroux and Boudeville, Friedmann defines his own sets of regions, called 'homogeneous' (analogous to Boudeville's use of the term) and 'interdependent'. The latter are close to the 'polarized regions' discussed above, being areas in which flows of goods and communications are predominantly towards a center, or core. For planning purposes, Friedmann divides up the periphery of the polarized region into several parts, to give the following:

|                       |   |              |
|-----------------------|---|--------------|
| core                  | } | 'center';    |
| upward transitional   |   |              |
| downward transitional |   |              |
| resource frontier     |   | 'periphery'. |
| special problem       |   |              |

The core region has the characteristics of the center but on a larger scale. Upward transitional regions are settled areas with growth potential, and like cores have net immigration. They are growing and have problems of capitalization. Downward transitional regions are old rural (or industrial?) economies in decline, whose resources suggest less intensive development than in the past, and where emigration is characteristic. Resource frontiers are zones of new settlement in which growth is potentially large in either agriculture or in mineral working associated with immigration and small new towns. Special problem regions are a category for areas that pose policy difficulties other than those above.

The main value of this simplified structure lies in three directions—its general applicability, its independence of scale, and its normative value. Being a generalized description, it has little explanatory value. In all the first three respects, however, it has certain advantages over the simple notion of a growth pole and growth center. In particular it deals with the whole of economic space, since the regions proposed are mutually exclusive and exhaustive. Moreover, as a normative framework it can distinguish between geographic areas in terms of necessary policies, while ensuring that regional problems are not dealt with in isolation, since solutions are possible only by considering the system as a whole.

Friedmann goes on to consider goals for spatial organization, and methods of implementation related to the regions specified above. The goals are related to the system as a whole and have two main requirements—that they must be adapted to the current phase of the system, and that they must be consistent with 'dominant

regional aspirations'. This distinguishes Friedmann from the majority of authors who are content to name criteria over which performance is to be optimized. He argues that one must use historically and *spatially* specific goals with elements outside the field of economics to create a reality in regional planning. By simulation, it would be possible to set up goals and to test criteria and policies by their relevance in achieving these goals.

The main issues of allocation of investments in space are discussed under the three polarities of:

growth versus welfare; imbalance versus balance; concentration versus dispersal. In many respects these are similar to each other, and raise the old issues of whether to emphasize growth at the expense of welfare and to concentrate investments in large agglomerations which will produce most multiplier effects, rather than scattering investment around in search of 'equity' or 'balance'. The choice in sectoral and social terms between balance and imbalance is expressed spatially in that between dispersal and concentration. However, the implication of such choices become clearer when seen against Friedmann's regions than when seen in Lloyd Rodwin's terms. For example, Rodwin (1961) advocates a policy of 'concentrated decentralization' when faced with the choices above. In this, small centers or cores are set up in the periphery, thus to some extent distributing investment but also taking some advantage of urbanization economies. This, however, is not a sufficient criterion for dealing with spatial investment problems, nor does it give much help when we are faced with different scales of problem. Obviously large scale concentrated decentralization (say from Paris to the metropolises) becomes merely 'centralization' if we look at the smaller region around a particular metropole, and so on. Friedmann's regions are a valuable substitute for this, since they allow the prescription of policies addressed to particular problems, in regions which are defined in relation to those problems. And because the polarized region can be thought of at any scale, the whole structure is not dependent on scale as is Rodwin's.

Goals for a society in the transitional phase are thus related to the removal of the periphery by substituting for it (dependent on the subregional structure above) a "single, interdependent system of urban regions" (Friedmann, 1966, p.54), and to the extension of a national system of factor and commodity markets. Friedmann concludes by specifying in great detail the kinds of action and policy that should be devised to deal with each of his types of region, sufficiently general to be applicable to any nation in the 'transitional' phase of development.

Friedmann thus presents us with a general model of the spatial aspects of regional economic growth, expressed chiefly in the descriptive and the normative, which goes further than the simple notions of growth pole and growth center. Moreover, his more recent work has attempted to expand the center-periphery formulation by making it applicable to all four phases of development, by considering explicitly cultural and political trends, and by extending the notion in an explanatory sense. The "General Theory of Polarized Development" (Friedmann, 1967) attempts to explain how the development process in its economic, social, and political aspects expresses itself in space, and how the resultant structuring of space changes through time.

#### 4 Evaluation

##### 4.1 *A spatial theory of growth?*

A theory may be thought of as a set of related definitions and hypotheses that attempts to explain a given phenomenon. Some of the hypotheses will be accepted on the basis of existing evidence, others will present a challenge to research. To what extent do growth pole and growth center notions constitute a 'theory' by this definition?

A theory of the spatial incidence of development—economic, and perhaps socio-cultural, and political—will have to relate a number of previously unrelated theories, hypotheses, and observed empirical regularities. The idea of a process of polarization taking place without reference to geographic space is a suitable starting point, even though it is confined, in the works reviewed above, to the economic field. However, in order to account for observed regularities, such as the appearance of a hierarchical structure of central places, we shall expect such a theory to embrace location economics and spatial organization theory, and in particular to account for strong tendencies towards agglomeration, which are apparent in society, possibly via theories of external economies. Moreover, we shall expect such a theory to account for these observed regularities, and hypothetical processes, both through space at one moment in time, and through time in a given space, such as a region or nation. We have seen in the literature various elements or components of such a theory presented in a number of unrelated, simplified, and non-rigorous formulations. We have seen attempts at using the concepts of growth pole and growth center in a normative sense, despite the lack of a full explanatory theory on which to base planning actions and policies. (The exigencies of the regional development problem are such that daily decisions must be made despite the lack of an adequate explanatory theory or model.) However, our expectations of a theory of the spatial incidence of growth would include the criterion that any explanations of existing distributions shall by the same formulation be applicable in a normative sense to the solution of planning problems. In other words, the explanatory sense of the theory should be sufficient to identify and quantify the specific areas in which the present structure of economic space falls short of that structure needed to implement the goals of the society under consideration, and to lead to the formulation of policies which will help the society to achieve the optimum distribution of population, industry, investment and urban equipment consistent with the achievement of its goals.

Clearly the notions of growth poles and growth centers reviewed do not constitute a 'theory' as defined above, although they do provide some basic elements of such a theory. We can, therefore, evaluate the literature in two main respects: firstly by reviewing the degree of rigor with which it deals with those areas that it purports to cover, and secondly by asking in what respects it falls short of our theoretical expectations, and indicating the three main directions in which research might proceed to fill in the gaps.

#### 4.2 *The positive contribution*

Most authors on growth poles agree that in order to act as a pole the industry under consideration must satisfy the three criteria of large size (and economic dominance), a rate of growth faster than that of the economy in which it is embedded, and a high degree of interlinkage with other sectors. By satisfying these criteria, it will thus be termed 'propulsive'. However, while this may be a sufficient condition to distinguish those sectors that will *transmit* growth (and this is doubtful), it says little of the way in which growth is *initiated*, and to this extent begs an important question. The process of polarization 'explains' the transmission of growth via interlinkages and external economies between one sector and all others, and, if the sector happens to be a fast growing one, the process would lead us to expect growth elsewhere in the economy. Hirschman's 'linkage effects' are a demonstration of this highly simplistic idea. It has been noted, however, that empirical demonstrations of the effects of polarization have, to date, not been able to distinguish between polarization effects and other effects, such as the influence of the market or the appearance of agglomeration economies. In other words, empirical demonstration has yet to show the strength of the polarization effect in relation to any of the other effects which can be claimed to

be 'growth inducing'. We must note that Isard's approach, using detailed input-output data in industrial complex analysis, is much closer to economic reality since it attempts to deal with many of the indirect interrelations between sectors, which the simple notions of growth poles and linkage do not, or can not.

Growth pole notions seem to stress the thesis of unbalanced growth developed by Hirschman. The prescription for progress which emerges from the pole notions is, however, somewhat undeveloped, since it relies simply on maximizing flows between sectors of the economy in the hope that this will generate multiplier effects which will, of course, be beneficial. However, this is not, nor can it be, related to specific policy objectives, and so its normative value is limited. We cannot agree, therefore, with Paelinck (1968), who maintains that the growth pole notion is "valuable chiefly to the extent that it clearly indicates the conditions under which accelerated regional development can occur". It does provide a set of conditions under which transmission of growth can be optimized, but these are by no means incontrovertible or exhaustive. Moreover, the concept says nothing about the initiation of growth and very little about the location and spatial distribution which might optimize growth.

What of the concepts which have been reviewed under the name 'growth center'? This concept specifically purports to deal with the distribution of growth, and allocation of investment, in real space. It has been shown that there have been few attempts to translate the ideas of propulsion, as developed with reference to firms in abstract space, into structures and distributions in geographic space; and indeed we have found this to be the major omission of the growth pole notions. Despite this, however, the growth center idea, particularly as generalized by Friedmann into the center-periphery structure, has great value. This lies in two main areas. Firstly, it provides what promises to be a dynamic theory of growth and development which can greatly enhance our static models, such as central place theory, and, as a corollary, it provides a basis for giving to development theory a spatial dimension. Secondly, growth center ideas are robust, if not rigorous; that is, they are addressed squarely to planning problems, and while they do not by any means answer these problems, they do provide the directions in which further research can proceed.

The explanatory value of the growth center idea is thus as yet limited and tends to be eclipsed by its normative value. Yet even the latter is more heuristic than specific. We gain from the literature at most certain normative preferences for action, such as: "it is better to concentrate investment in centers, than scatter it around", "bigger centers will be better than smaller ones in the amount of growth produced from a given level of investment", "there are a number of sub-optimum structures for generating and transmitting growth at all phases in the development of a polarized region—and the central place hierarchy should be modified to this extent", and "development, in spatial terms, will consist of promoting integration of the periphery, by a single interdependent system of cities".

One of the most valuable aspects of the idea of the growth center is that it gets away from the reliance of growth poles on the big firm or big industry as a basis for growth. Several authors have stressed the importance of the tertiary sector, and of the existence in an agglomeration of many smaller firms providing the specialized services on which the external economies of agglomerations are hypothesized to be based.

Attempts have also been made to generalize the notion of the center into the socio-cultural and political fields. While these attempts have not yet produced results, their existence is encouraging as representation of the awareness now common that explanations of and prescriptions for growth and development in spatial terms must cover many non-economic variables if they are to claim success.



Perhaps the most satisfactory explanations in the literature are those developed by Friedmann, directed at the spatial incidence of growth as related to the stages of national development. This is one of the criteria mentioned above by which we would recognize a 'theory'. Friedmann's center-periphery structure is, it is true, a highly generalized one which does not approach specific cases, other than that of Venezuela. However, we would defend this quest for general applicability, even at the expense of rigor, as being most appropriate and relevant to the development of a satisfactory theory. The 'style' of growth center literature has necessarily been somewhat generalized, since it seeks to incorporate variables which are difficult to conceptualize and quantify.

The most serious omissions in the literature concern the absence of explicit statements about the relationship between polarization and empirically observed regularities on the ground, and the inadequate treatment of the whole of external economies and the pronounced tendencies towards agglomeration. These are the main areas in which growth pole and growth center notions fall short of our expectations in terms of a theory of the spatial incidence of economic growth. A brief discussion of them can, therefore, provide a suitable conclusion to this review by pointing in significant directions for further research.

#### 4.3 *The major deficiencies*

4.3.1 *Agglomeration and externalities.* There is general agreement that growth center and growth pole notions will prove of use only if they treat adequately the tendency of economic, social and political processes to produce agglomerated phenomena. Much has been written of the role of external economies in the growth of urban areas and centers, and most of this has been on the basis of economic variables. It is not the purpose of this paper to review the literature on externalities, but some of the distinctions and contradictions within it must be established in order to view more clearly the relationship between agglomeration and the processes of polarization.

Most authors appeal to 'external economies' in order to 'explain' agglomeration. "... any adequate treatment of this phenomenon (of polarization) should take account of the pronounced tendency for industrial growth to be oriented primarily towards industrial areas because of the external economies which the latter generate ...." (Hansen, 1967, p.718).

It seems intuitively obvious that cities or central places can, by their agglomeration, provide certain advantages which are not to be found either in smaller agglomerations or in dispersed distributions.

In his classical formulation, Hoover (1968) discusses the three types of external economies which he holds to be important in the growth of large agglomerations. The first, the 'principle of multiples', refers to the possibility of increased specialization by firms. Certain operations that would normally have to be carried out by a firm can be contracted to other firms specializing in such operations, and this at lower cost. The second principle, 'massing of reserves', means that in a large city firms can carry proportionally smaller stocks of materials than they can in an isolated location, since they are able to depend on their ability to secure more at short notice. The third principle is that of 'bulk transactions', referring to the economies of large-scale transfer, and the reduction in unit cost which occurs when many firms are consuming large quantities of basic materials such as electricity.

Alonso (1968), discussing industrial location in relation to economic development, adds more 'economies' to this growing list. He points out that in transitional societies, conditions in the large agglomeration tend to be better known than do conditions in the rural periphery, and that businessmen making investment decisions

are attracted to the place in which uncertainty is minimized. This is a rational basis for investment decision, and is comparable to Hirschman's homily in explaining the continued growth of agglomerations—"nothing succeeds like success". Hirschman, however, claims (unlike Alonso) that this is an irrational decision and urges investors to look to the periphery. Other advantages of agglomeration in transitional societies discussed by Alonso include the need for personal contact in a society in which communication systems are scarce or unreliable, and the importance in these societies of the ritual of social contact, much of which has been replaced by standardized methods of business in developed economies. All this pull in favor of the big city is magnified by the concentration there of entrepreneurial and managerial resources. This tends to emphasize the distinction between center and periphery and to underline the appearance of 'primacy'—the domination of a single, big city in economies which are going through the transitional stage of industrialization, followed by a decline of primacy in favor of a more fully developed central-place system, manifesting the increasing integration of economic space.

Von Boverter (1964) points to other forces in agglomeration. Because of higher costs, nominal wages are also higher in the urban area than outside it. This offsets the advantages of the agglomeration. However, as labor mobility grows and agglomeration economies increase, the urban areas will still tend to have a balanced advantage, though they may become suburbanized. Friedmann (1955), however, claims that as development occurs there is a shift from labor and material-oriented industry to market-oriented industry and this confirms the trend towards agglomeration. Friedmann could not find evidence to support this, however, since *all* types of industry tended to be strongly oriented toward the agglomerations (in the T.V.A. study).

Both Von Boverter (1964) and Marcus (1965) have looked at agglomeration economies in two subsets—localization economies, which occur when several firms within an industry are agglomerated, and urbanization economies, which result from the general agglomeration of labor and market and operate in different industries. Marcus, looking at New Jersey, shows that a number of industries are growing faster than one would expect from their national growth rates, weighted by the growth of population in the area. He attributes the 'extra' growth to the two sets of economies mentioned.

From the above, it seems that there is no clear agreement on what constitutes external economies or agglomeration economies and, even worse, the various ideas are often non-comparable. Further, Speigelman (U.S. Department of Agriculture, 1966, p.13) claims that although agglomeration and external economies appear to be correlated in their spatial incidence, there "is almost no quantitative evidence as to the role of external economies in the agglomerating process". This is undoubtedly true, for since there is almost no agreement on how external economies are defined, there is little hope of quantification.

For growth center notions, this lack of evidence is serious since it means that while we observe that cities grow, and because of their growth attract more firms, investment and people, and while we have some idea about how this growth is transmitted, we have little evidence except *post facto* to indicate how growth is best initiated, nor even why some agglomerations grow faster and larger than others. We therefore have no firm theoretical grounds for planning growth at particular locations. Despite this, growth center notions have proliferated, based on two concepts—the growth pole (industry) discussed above, and, the only really quantified aspect of external economies, the number and kinds of services and functions performed by cities of various sizes.

From central place studies, theoretical and empirical, we have a foundation on which to discuss the ability of different locations to provide local services, which

themselves can operate as growth generators. It is no accident that much of the growth center work incorporated in the fifth French plan concentrates on the analysis and development of tertiary services.

*4.3.2 Size and scale of growth centers.* There is also little agreement on the size of growth centers, and in the planning field, on the optimum size, given a set of goals and constraints. Clearly, if a theory of the spatial incidence of growth is to be developed, it must include some postulates about the size of growth centers, and the relationship between size and rate of growth, at a given state of development and in a given socio-cultural system. Unfortunately, this is very close to the issue of the optimum size of a city, which is a problem replete with hypotheses, many of them ill-founded, and notably lacking in evidence.

Advocates have claimed sizes between 10 000 and 1 000 000 to be optimum, but in general they consider only the cost aspects of size and ignore the fact that we are dealing with an urban system of a hierarchical nature in which cities are of very different sizes, so that a single optimum size is infeasible in any case, unless the whole system is somehow changed. Most authors writing on size have some intuitive notion that average and marginal costs will at first fall and then rise as a center grows in size, appealing to the costs of congestion and of commuting in 'evidence'. However, as Alonso points out, the size of a city must be considered also in relation to its productivity, measured say in terms of disposable income per head, from which costs per head can be subtracted. This produces a more realistic concept of the 'optimum', since productivity and costs will vary presumably in some consistent fashion with size. However, as Alonso (1966, p.8) puts it, despite many assertions by many authors: "the fact of the matter is that there is no reliable knowledge of relevant urban sizes". Certainly there is no evidence that there exists a city size beyond which marginal costs outrun marginal productivity, and in this respect there are no indications that an agglomeration is 'too big', despite the attempts being made in Western European countries to decentralize their major capital cities. Indeed, there is accumulating a body of evidence to suggest that per capita income, productivity in manufacturing, wholesale sales per employee, and some other measures, all continue to rise, without apparent limit, with increase in the size of agglomeration (measured, admittedly, in a cross-sectional sense).

Again, this paucity of evidence is somewhat serious for growth center concepts, since there appear to be no guidelines as to optimum size or even as to threshold size for growth to be self-sustaining, and rarely in the literature have attempts been made to relate such measures of size to the state of economic development under consideration.

*4.3.3 Growth centers, and the central-place hierarchy, through time.* A final major area of deficiency, and therefore a topic of future research, is the relationship of growth center and growth pole concepts to the observed empirical regularities of the central-place type, and to the phenomenon of primacy in countries of the developing world.

In the survey of the literature, there are noticeable areas of compatibility between some authors. In particular, Berry (1964, p.129) stresses the relationship between the development of the central-place system and a state of entropy in the socio-economic system, "achieved in the steady state of a stochastic process . . . at its maximum if this process is unconstrained", and "if the rank size rule for cities obtains". Hirschman (1958) hypothesizes, and to a limited extent demonstrates, that a major characteristic of differences between 'stages' of economic development is the degree or complexity of interlinkages between sectors of the economy. As the nation or region becomes more developed, the interlinkages and interdependencies are maximized. Similarly, Friedmann (1966) sees the spatial aspect of economic

development as one in which the region under consideration progressively replaces its center-periphery structure with a single system of cities extending throughout the economic space under consideration.

All this suggests therefore that 'primacy' is somehow related to the 'phase of development' and that the central-place structure is a spatial state manifesting the achievement of an equilibrium in socio-economic development. Fortunately, early conclusions claiming to demonstrate the invalidity of this relationship, notably by Berry, have themselves been shown to be invalid, and in fact, while primacy is very rare in truly underdeveloped countries, it begins to appear in the 'take-off' stage, and then declines as further development takes place (El Shaks, 1965). Thus, primacy is a normal aspect of the early stages of development characterized by a state of negentropy, and is corrected by the progress of the system towards its entropic, complex state of equilibrium.

Growth center concepts must be related to the state of this system if they are to be developed into a theory. Is the notion of a center and periphery merely another way of describing a negentropic state of primacy? Will the development of inter-sectoral linkages between industries encourage, therefore, the development of a central-place system which will integrate the periphery? Is there an optimum central-place structure at a given level of development? Is such a structure in some way a cause or an effect of the degree of interlinkage between sectors? All these questions are begged by the growth center concept, and imply that there will be different rates of growth and different optimum sizes of center at different stages of development, and that their optimum locations will vary due to this.

## 5 Conclusion

Precise conclusions are difficult to draw from a field so ill-defined and confused as the one reviewed in this paper. Nonetheless, certain prevailing characteristics can be pointed out.

Firstly, it is evident that the explanatory value of the growth pole and growth center notions is limited. This is undoubtedly because both notions deal with only a limited concept which is part of a much more complex system described more realistically by the detailed input-output table on the one hand and by the notions of the central-place system on the other. Growth pole discussions have tended to be somewhat myopic, concentrating on the direct links between a hypothetical industry and a few others in an economy assumed to be closed, to the total neglect of the enormous amount of background variation and indirect linkage taking place. Similarly, growth center notions tend to concentrate on a particular aspect of the central-place system and attempt to treat it in isolation. It is for this reason that the extraordinary gap discovered in the literature—namely, the lack of any satisfactory explanation of how the existence of a growth pole and the process of polarization in economic space appear on the ground in terms of the distribution of industries and agglomerations—is so obvious.

The hypothesized process of polarization has heuristic attraction but, again, because of the over-simplified framework in which it is discussed, its value as an explanatory model is not great. The absence of any development of the idea into a set of precisely related equations linking sectoral to spatial development, and of any empirical study which separates the process of polarization from other effects, is noteworthy in this respect, and the process thus begs many more questions than it answers.

As a description of the realities of the occurrence and spatial distribution of development, economic and social, the notions of pole and center are of limited help. We cannot agree that they form in any way a conditional theory of economic growth,

since the conditions hypothesized are insufficient to distinguish a growth from a non-growth situation, and the criteria developed are inadequate to distinguish between a growth center and a non-growth center, either in the present or in the (normative) future. Moreover, the notions do not treat adequately either the influence of external economies or the phenomenon of agglomerations.

It is in the normative that the notions have their greatest contribution to make. The center-periphery concept in particular, as described by Friedmann and discussed by others, is the most promising direction reviewed. By dealing with the whole of economic space in a given region, rather than particular points or areas of it, and by defining sub-regions of the periphery in terms of the problems for which solutions are sought, it is a valuable step towards the prescription of policies for the distribution of economic and social development given a set of goals. The links developed between the model and the empirical regularities observed at different states of economic development are attractive as a theory of the spatial distribution of development through time. Attempts to consider other than economic variables in this context are moves in the right direction.

The value of the center-periphery model in particular cases is limited as yet by its relative lack of development. Nonetheless, by treating the system as a system rather than by picking up and isolating parts of it, it will be a valuable tool in regional planning. The French notion of the 'armature urbaine', which shows features in common with the center-periphery concept, is being put into practice through the regionalization of the French budget. The effect on growth, both regional and national, of concentrating investment in certain sectors of the economy, at the eight back by which the notion can be evaluated. The development of a series of regional statistics by the French government should make such an evaluation a possibility in the near future.

#### References

- Alonso, W., 1964, "The form of cities in developing countries", *Papers and Proceedings of the Regional Science Association*, XIII.
- Alonso, W., 1966, *Location, Primacy and Economic Development*, Center for Planning and Development Research, Discussion Paper, mimeo., Berkeley, California, Center for Planning and Development Research, University of California, August.
- Alonso, W., 1968a, *Industrial Location and Regional Policy in Economic Development*, Center for Planning and Development Research, Working Paper 74, Berkeley, California, Center for Planning and Development Research, University of California, February.
- Alonso, W., 1968b, *Equity and Its Relation to Efficiency in Urbanization*, Center for Planning and Development Research, Working Paper 78, Berkeley, California, Center for Planning and Development Research, University of California, July.
- Antoine, J. C., 1962, "Recherches statistiques sur la structure des agglomérations", *Cahiers de l'I. S. E. A.*, L, 11.
- Antoine, S. and Weill, G., 1968, "Les métropoles et leur région", *L'Espace et les Pôles de Croissance*, Ed. J. R. Boudeville, Biblio. de L'Economie Contemporaine (Presse Universitaire de France, Paris).
- Athanasopoulos, D., 1964, "Développement régional et aménagement du territoire en Grèce", *Revue d'Economie Politique*, 1, 74.
- Audouin, 1962, "Etude du coût urbain appliqué à une ville de moyenne importance", *Ecole Nationale d'Administration*.
- Aujac, H., 1960, "La hiérarchie des industries dans un tableau des échanges industriels et ses conséquences dans la mise en oeuvre d'un plan national décentralisé", *Revue Economique*, XL, May.
- Aydalet, P., 1965a, "Etudes sur les processus de polarisation et sur les réactions des industries anciennes à la lumière de l'expérience à Lacq", *Cahiers de l'I.S.E.A.*, L, March.
- Aydalet, P., 1965b, "Le rôle économique des autoroutes", *Revue d'Economie Politique*.
- Aydalet, P., 1965c, "Note sur les économies externes et quelques notions connexes", *Revue Economique*, November.

- Bairau, P., 1955, "Le coût de l'agglomération troyenne", Ecole Nationale d'Administration.
- Ballargeon, J.-P., 1961, "Le rôle des pôles dans le développement. Exposé introductif", *Développement et Civilisation*, 5, January-March.
- Bauchet, P., 1955, *Les Tableaux Economiques, Analyse de la Région Lorraine* (Genin, Paris).
- Bauchet, P., 1961a, "La compatibilité économique régionale et son usage", *Economie Appliquée*, XIV, January.
- Bauchet, P., 1961b, "La programmation régionale", *Théorie et Politique de l'Expansion Régionale*, Institut de Science Economique, Université de Liège.
- Bauchet, P., 1965, "Regional development possibilities in France", *Area Redevelopment Policies in Britain and the Countries of the Common Market*, U. S. Department of Commerce, Area Redevelopment Administration, Washington, D. C.
- Bauchet, P., 1966, *La Planification Française* (du Seuil, Paris).
- Beguín, H., 1963, "Aspects géographiques de la Polarisation", *Revue Tiers-Monde*.
- Berry, B. J. L., 1964, "Cities as systems within systems of cities", *Regional Development and Planning*, Eds. W. Alonso and J. Friedmann (M. I. T. Press, Cambridge).
- Berry, B. J. L., 1961, "City size distribution and economic development", *Economic Development and Cultural Change*, 9, July.
- Bettelheim, C., 1966, *Problèmes Théoriques et Pratiques de la Planification*, Paris.
- Bloch-Laine, F., 1962, "Sept ans d'incitation à l'expansion régionale: bilan et leçons", *Revue Juridique et Economique du Sud-ouest*, 9.
- Boudeville, J.-R., 1957, "Contribution à l'étude des pôles de croissance Brésiliennes: une industrie motrice, la sidérurgie du Minas Gerais", *Cahiers de l'I. S. E. A.*, F, 10.
- Boudeville, J.-R., 1958, "L'économie régionale, espace opérationnel", *Cahiers de l'I. S. E. A.*, L, 3.
- Boudeville, J.-R., 1959, "La région plan", *Cahiers de l'I. S. E. A.*, L, 6.
- Boudeville, J.-R., 1960, "L'espace opérationnel macroéconomique, la région plan", *Cahiers de l'I. S. E. A.*, L, 6.
- Boudeville, J.-R., 1961a, *Les Espaces Economiques*, Paris: Que sais-je? 950 (Presse Universitaire de France, Paris).
- Boudeville, J.-R., 1961b, "Les zones industrielles et l'utilisation du territoire", *Cahiers de l'I. S. E. A.*, L, 9.
- Boudeville, J.-R., 1961c, "Un modèle des mouvements commerciaux interrégionaux en France", *Cahiers de l'I. S. E. A.*, L, 9.
- Boudeville, J.-R., 1964a, "Hiérarchie urbaine et aménagement des villes", *Revue d'Economie Politique*, 74.
- Boudeville, J.-R., 1964b, "Les instruments de la région de programme", *Cahiers de l'I. S. E. A.*, L, 14.
- Boudeville, J.-R., 1964c, "Note sur l'intégration des espaces économiques", *Cahiers de l'I. S. E. A.*, L, September.
- Boudeville, J.-R., 1966, *Problems of Regional Economic Planning* (Edinburgh University Press).
- Boudeville, J.-R., 1968a, *L'Espace et les Pôles de Croissance*, Paris: Biblio. d'Economie Contemporaine (Presse Universitaire de France, Paris).
- Boudeville, J.-R., 1968b, "Les notions d'espace et d'intégration", *L'Espace et les Pôles de Croissance*, Ed. J.-R. Boudeville (Presse Universitaire de France, Paris).
- Boudeville, J.-R., undated, *Les Programmes Economiques*, Paris: Que sais-je? 1073 (Presse Universitaire de France, Paris).
- Bourguinat, H., 1961, "Espace économique et intégration européenne", *Société d'Etudes pour le Développement Economique et Social*, Paris.
- Bourguinat, H., 1964, "Economies et déséconomies externes", *Revue Economique*, June.
- Boventer, E. von, 1964, "Spatial organization theory as a basis for regional planning", *Journal of the American Institute of Planners*, May.
- Brounin, A., 1957, "La notion de coût des agglomérations applicable à huit villes de Seine-et-Oise", Ecole Nationale d'Administration.
- Buchanan, J., and Stubblebine, W. C., 1962, "Externality", *Economica*, November.
- Cazes, B., 1962, *La Planification en France et le IV<sup>ème</sup> Plan* (Lepargne Paris; Paris).
- Centre des Etudes et Recherches Economiques et Sociales, 1959, "Niveaux optima des villes", Faculté de Droit, Université de Lille, July.
- Chaineau, A., 1965, "Un modèle d'analyse de l'espace économique Français", *Revue d'Economie Politique*, 1, 75, January-February.
- Chinitz, B., 1961, "Contrasts in agglomeration, New York and Pittsburg", *American Economic Review*, May.
- Chisholm, M., 1964, *Rural Settlement and Land Use*, (Hutchinson, London).

- Clark, C., 1945, "The economic functions of a city in relation to its size", *Econometrica*, 13, 2, April.
- Commissariat Général du Plan, 1966, *V-e Plan de Développement Economique et Social*, Paris.
- Council of Planning Librarians, 1968, *Concept of an Optimum City Size, a Selected Bibliography*, Exchange Bibliography, 52, May.
- Davin, L. E., 1961, "Les conditions de croissance des économies régionales dans les pays développés", *Théorie et Politique de l'Expansion Régionale*, Institut de Science Economique, Université de Liège.
- Davin, L. E., 1964a, *Economie Régionale et Croissance* (Genin, Paris).
- Davin, L. E., 1964b, "La politique Belge de développement régionale", *Revue d'Economie Politique*, 1, 74.
- Davin, L. E., 1965, "Phénomènes de polarisation; seuils de croissance et expansion économique en Afrique Centrale", *Cahiers de l'I. S. E. A.*
- Davin, L. E., Degeer, L., and Paelinck, J., 1959, *Dynamique Economique de la Région Liégeoise* (Presse Universitaire de France, Paris).
- Davis, O. A., and Winston, A., 1962, "Externalities, welfare and the theory of games", *Journal of Political Economy*, June.
- Delaman, 1956, "Contribution à l'étude des coûts des agglomérations, Vitry-le-François", Ecole Nationale d'Administration.
- Della Porta, P. G., 1960, "La planification nationale et régionale en Italie", *Economie Appliquée*, 13, 4.
- Derwa, L., 1957, "Analyse input-output de la région Liégeoise", *Revue de Conseil Economique Wallon*, September-November.
- Duncan, O. D., 1951, "The optimum size of cities", in *Reader in Urban Sociology*, Eds. P. K. Hatt and A. J. Rice (Glencoe, Illinois).
- Echard, J., 1961, "La collaboration des initiatives publiques et privées pour le développement d'un pôle industriel: Lacq", *E. E. C. Conference sur les Economies Régionales*, Brussels.
- El Shaks, S., 1965, "Development, primacy and the structure of cities", unpublished Ph. D. Dissertation, Harvard University.
- Flamont, M., 1964, "Concept et usage des économies externes", *Revue d'Economie Politique*, 1, June.
- Fleming, M., 1955, "External economies and the doctrine of balanced growth", *Economic Journal*, June.
- Fogarty, M., 1964, "Quelques leçons de la politique Britannique de localisation de l'industrie", *Revue d'Economie Politique*, 1, 74.
- Fox, C., 1966, *The Role of Growth Centers in Regional Economic Development*, Department of Economics, State University of Science and Technology, Ames, Iowa, September.
- Friedmann, J. R., 1955, *The Spatial Structure of Economic Development in the Tennessee Valley* (University of Chicago Press, Chicago).
- Friedmann, J. R., 1964, "Regional development in a post-industrial society", *Journal of the American Institute of Planners*, 30, 2, May.
- Friedmann, J. R., 1966, *Regional Development Policy—A Case Study of Venezuela* (M. I. T. Press, Cambridge).
- Friedmann, J. R., 1967a, "A general theory of polarized development", mimeo., Santiago, Chile, August.
- Friedmann, J. R., 1967b, "Regional planning and nation building—an agenda for international research", *Economic Development and Cultural Change*, 16, 1, October.
- Friedmann, J. R., 1968 "The role of cities in national development", mimeo., Santiago, Chile, February.
- Gottlieb, M., 1945, "The theory of optimum population in a closed economy", *Journal of Political Economics*, L, III, December.
- Gravier, J. F., 1964, *L'Aménagement du Territoire et l'Avenir des Régions Françaises* (Flammarion, Paris).
- Guichard, O., 1965, *Aménager la France* (Laffont Gonthier, Paris).
- Hansen, N. M., 1965, "Unbalanced growth and regional development", *Western Economic Journal*, IV, 1.
- Hansen, N. M., 1967a, "Development pole theory in a regional context", *Kyklos*, XX.
- Hansen, N. M., 1967b, "Human resources and economic development: some lessons from French experience", *Southern Economic Journal*, XXXIV, 1, July.
- Hansen, N. M., 1968 "The role of growth centers in comprehensive regional development policy", mimeo., Paper prepared for the Economic Development Administration of the U. S. Department of Commerce, Washington, D. C., February.

- Hansen, N. M., 1968, *French Regional Planning* (Indiana University Press, Bloomington, Indiana), May.
- Hackett, J., and Hackett, A. M., 1963, *Economic Planning in France* (Harvard University Press, Cambridge).
- Hautreux, J., 1966, "Le rôle des métropoles d'équilibre dans l'armature urbaine", *Revue Juridique du Sud-Ouest Sér Ec.*
- Hermida, H. R., 1960, "La programmation régionale en Espagne", *Economie Appliquée*, 13, 4.
- Hirschman, A. O., 1958, *The Strategy of Economic Development* (Yale University Press, New Haven).
- Hoeffler, J., 1962, "L'aménagement régional de la Belgique", *Cahiers d'Urbanisme*, 42 and 43.
- Hoover, E. M., 1968, *The Location of Economic Activity* (McGraw-Hill, New York).
- Ilchman, W. F., and Barghava, R. C., 1966, "Balanced thought and economic growth", *Economic Development and Cultural Change*, July.
- Institut des Civilisations Différentes Bruxelles, 1952, *L'Attraction Exercée par les Centres Urbains Industriels dans les Pays en Voie d'Industrialisation*, Brussels.
- Institut de Science Economique de l'Université de Liège, 1961, *Théorie et Politique de l'Expansion Régionale*, Liège.
- International Bank for Reconstruction and Development, 1968, *French Academic and Private Research on Economic Development Problems*, mimeo., February.
- Isard, W., 1960, "Industrial complex analysis", in *Methods of Regional Analysis* (M. I. T. Press, Cambridge).
- Isard, W., and Schooler, E. W., 1959, "Industrial complex analysis: agglomeration, economics and regional development", *Journal of Regional Science*, Spring.
- Jeanneney, J.-M., 1956, "Principes pour une politique nationale de développement des économies régionales", *Revue Economique*, 6, November.
- Korner, H., 1967, "Industrielle Entwicklungspole als Instrument der Regional Politik in Entwicklungslandern", *Kyklos*, XX.
- Labasse, J., 1968, "Le rôle des équipements tertiaires supérieurs dans la polarisation de la vie régionale en Europe occidentale" *L'Espace et les Pôles de Croissance*, Ed. J.-R. Boudeville, Biblio: d'Economie Contemporaine (Presses Universitaires de France, Paris).
- La Documentation Française, "Problèmes économiques, dix ans de décentralisation industrielle", *La Documentation Française*, 945.
- La Documentation Française, "Problèmes économiques, les grands aménagements régionaux", *La Documentation Française*, 957.
- Lagujie, J., 1964a, "Aménagement du territoire et développement économique régionale en France 1945-1964", *Revue d'Economie Politique*, 1, 74.
- Lagujie, J., 1964b, "Développement économique régional et aménagement du territoire", *Revue d'Economie Politique*.
- Lagujie, J., 1966, "La planification régionale en France", *Revue Juridique et Economique du Sud-Ouest*.
- Laure, A., 1961, "Les programmes de modernisation et d'équipement des agglomérations", *Moniteur des Travaux Publics et du Bâtiment*, 15, 9.
- Lebret, R. P., 1961, "Agglomérations et pôles de développement", *Cahiers d'Urbanisme*, 33.
- Le Pas, J., 1966, "Essai de typologie des économies externes urbaines", *Cahiers de l'I. S. E. A.*, 2, 6, June.
- Marcus, M., 1965, "Agglomeration economics: a suggested approach", *Land Economics*, 41, August.
- Martin, J., 1962, "Bibliographie de science économique régionale", *Cahiers de l'I. S. E. A.*, 1, 10, May.
- Martin, R. J., 1966, *The Correlates and Significance of Settlement Size*, Research Study 1, School of Town and Country Planning, Edinburgh, November.
- Martin Lobo, M., 1964, "La politique de développement régionale en Espagne", *Revue d'Economie Politique*, 1, 74.
- Massé, P., 1964, "L'aménagement du territoire, projection géographique de la société de l'avenir", *Revue d'Economie Politique*, 1, 74.
- Meade, J. E., 1952, "External economies and diseconomies in a competitive situation", *Economic Journal*, LXII.
- Ministère de la Construction, 1958, 1960 supplement, *Bibliographie de l'Aménagement du Territoire National et Régional*, Paris.
- Myrdal, G., 1957, *Economic Theory in Underdeveloped Regions*, London.
- Neutze, G. M., 1965, *Economic Policy and the Size of Cities* (Australian National University, Canberra).



- Paelinck, J., 1959, "Possibilisme et pôles de croissance", *Economie Appliquée*, 1 and 2.
- Paelinck, J., 1965, "La théorie du développement régionale polarisé", *Cahiers de l'I. S. E. A.*, L, 15.
- Paelinck, J., 1968, "Systématisation de la théorie du développement régionale polarisé", *L'Espace et les Pôles de Croissance*, Ed. J-R. Boudeville (Presse Universitaire de France, Paris).
- Paelinck, J., and Waelbroeck, J., 1963, *Programmation Economique et Modèles Econométriques de Croissance* (Thone, Liège).
- Penouil, M., 1964, "La région dans la planification économique des pays en voie de développement", *Revue d'Economie Politique*, 1, 74.
- Perroux, F., 1950, "Economic space, theory and applications", *Quart. J. Econ.*, LXIV.
- Perroux, F., 1952, "Les pôles de développement", *Economie Appliquée*.
- Perroux, F., 1955, "Note sur la notion des pôles de croissance", *Economie Appliquée*, 1 and 2.
- Perroux, F., 1961, "La firme motrice dans une région et la région motrice", *Cahiers de l'I. S. E. A.*, AD, March.
- Perroux, F., 1962, "Le IV-e plan Français", Que sais-je? 1021, Paris.
- Perroux, F., 1964a, "La notion du développement", *L'Economie de XX Siècle* (Presse Universitaire de France, Paris), p.155.
- Perroux, F., 1964b, *L'Economie de XX Siècle* (Presse Universitaire de France, Paris).
- Perroux, F., 1964c, "Les points du développement et les foyers de progrès", *L'Economie de XX Siècle* (Presse Universitaire de France, Paris), p.192.
- Perroux, F., 1964d, "Les pôles du développement et l'économie internationale", *L'Economie de XX Siècle* (Presse Universitaire de France, Paris), p.172.
- Perroux, F., 1965, *Technique Quantitative de Planification*, (Presse Universitaire de France, Paris).
- Perroux, F., 1968, "Les investissements multinationaux et l'analyse des pôles de développement d'intégration", *Revue Tiers-Monde*, IX, 34.
- Perroux, F., undated, "Matériaux pour une analyse de la croissance économique", *Cahiers de l'I. S. E. A.*, D, 9.
- Pinchemel, P., 1966, "Réflexions sur les relations entre l'industrie et la ville", *Cahiers de l'I. S. E. A.*, 2, 6, June.
- Poittier, P., 1963, "Axes de communication et théorie de développement", *Revue Economique*, XIV, January.
- Political and Economic Planning, 1961, *Economic Planning in France*, Political and Economic Planning, Paris, XXVII, 454.
- Political and Economic Planning, 1963, *French Planning—Some Lessons for Britain*, Political and Economic Planning, XXIX, 475, September.
- Ponsard, C., 1955, *Economie et Espace* (SEDES, Paris).
- Premier Ministre et Commissariat du Plan, 1966, 1967, *Projet de Loi de Finances, 1966 and 1967* (see especially Rapport sur Régionalisation du Budget, 1966 and 1967), Paris.
- Revue d'Economie Politique, 1964 (issue devoted to regional planning problems, of which some related to the Growth Pole Concept), LXXIV, No.1, 74.
- Rodwin, L., 1961, "Choosing regions for development", *Public Policy*, Eds. Freidrich and Harris, XII, January.
- Rosenfeld, F., 1962, "Les firmes motrices et la comptabilité régionale", *Cahiers de l'I. S. E. A.*, L, 11.
- Rosenfeld, F., 1964, "Structure et perspectives économiques de la Province de Turin", *Metra*, III, 4.
- Rottier, P., 1960, "Développement économique et équipement urbain", *Annales du CREDOC*, March.
- Scitovsky, T., 1954, "Two concepts of external economies", *Journal of Political Economics*, April.
- Shindman, B., 1955, "An optimum size for cities", *Canadian Geographer*, 5.
- SODIC—Société pour la Conversion et le Développement Industriel, 1966, *Guide d'Implantation Industrielle dans les Régions de l'Ouest de la France*, Paris.
- Thomas, G., 1965, "Le coût des différents types d'agglomération", Ecole Nationale d'Administration.
- Tiebout, C., 1960, "Economies of scale and the metropolitan governments", *Review of Economics and Statistics*.
- Tinbergen, J., 1964, "Un modèle de la dispersion géographique de l'activité économique", *Revue d'Economie Politique*, 1, 74.
- U. S. Department of Commerce, Area Redevelopment Administration, 1965, *Area Redevelopment Policies in Britain and the Countries of the Common Market*, Washington.
- U. S. Department of Agriculture: Economic Research Service, 1966, *Analysis of Urban Agglomeration*, Agricultural Economics Report, 96, June.

- Van Os, F. J. J. H. M., 1960, "La politique de l'industrialisation régionale aux Pays-Bas", *Economie Appliquée*, 13, 4.
- Vandeputte, R., 1960, "La politique d'expansion régionale en Belgique", *Economie Appliquée*, 13, 4.
- Viner, J., 1958, "Stabilité et progrès, les problèmes de la Pauvreté", *Economie Appliquée*, 1.
- Wellisz, S., 1964, "On external diseconomies and the government assisted invisible hand", *Economica*.
- Whittlesey, D., 1954, "The regional concept and the regional method", in *American Geography: Inventory and Prospect*, Eds. P. James and C. Jones (Syracuse University Press, New York).
- Widmer, G., 1953, "L'inégalité dans la grandeur des villes; des corrélations économiques", *Revue Economique*, 3.

Note: The works cited in this bibliography include only those which deal specifically with growth pole or growth center concepts, external economies, agglomeration, optimum size, and certain aspects of regional development. Except for certain less well known French sources, works dealing generally with regional planning have been excluded.

## 26 Criteria for a Growth Centre Policy

Niles M. Hansen

### INTRODUCTION

Growth centre strategies have been instituted in numerous industrial countries during the past two decades. For the most part they have been designed to promote the development of lagging regions by concentrating investment so as to reap scale and agglomeration economies. In addition, growth centres also have been conceived as points of attraction for migrants who otherwise would go to large congested urban areas.

To date, these strategies have not been notably successful, primarily because they have suffered from a tendency towards proliferation of a large number of relatively small centres. Even where a few large centres have been selected, as in the case of France's eight *métropoles d'équilibre*, inadequate funding has been a problem.

Although there is growing realization that a viable growth centre strategy demands greater selectivity in choice of centres, it still is uncertain that the spread effects from induced growth centres can really lift income and employment opportunity levels in lagging or declining regions to levels comparable to those in more advanced regions. Indeed, it is possible that the public works bias which inevitably accompanies growth centre policies has done a disservice by shifting attention away from the critical health, education, and social problems which plague underdeveloped regions. Preoccupation with spread effects also has hindered efforts to explore the opportunities which growth centres in intermediate areas (that is, viable growth centres which are neither merely relatively bright spots in lagging areas nor large congested cities) offer to migrants from lagging regions.

Growth centre theory in its present state does not provide specific criteria for identifying the location of relevant urban centres, how big they should be, or what kinds of investments should be placed in them.

Reprinted from *Growth Poles and Growth Centres in Regional Planning*, ed. Antoni Kuklinski (The Hague: Mouton, 1972), pp. 103-124, with permission of the author and the publisher. © 1972 by Mouton & Co.

This is true of the French School as well as later formulations, and it is especially true for the problems of lagging areas in industrialized countries. Nevertheless, growth centre analysis implicitly or explicitly involves deliberate policies; otherwise one is left with merely a description of the spatial distribution and growth of population and economic activity.

This paper develops the general outline of a growth centre policy in the context of the United States. However, before examining the central issues in our approach it may be useful to specify certain notions that it either ignores or rejects. These include the doctrine that economic growth should be 'balanced' geographically, the approach which takes growth centres to be either exclusively or primarily generators of spread effects to hinterlands, reliance on central place theory in growth centre determination, and the identification or association of new towns with growth centres.

#### 1. WHAT IS NOT AN ISSUE

A common theme in growth centre strategies is the notion of 'balanced' growth or a 'balanced' distribution of economic opportunity. However, it is rarely specified in more concrete terms what this means. Does it mean that equality of *per capita* public overhead capital, income, or economic activity (however defined) should be the goal? Should the growth of less developed regions be promoted solely by moving resources to them or creating new resources within their boundaries? Precisely what public and private activities should be located in various types of cities and regions? What effects will the location of various types of activities in a given region have on other regions as a result of induced activities (on both the supply and demand sides) of an inter-regional nature? What conflicts might arise between maximizing regional and national welfare and how should they be resolved? It would seem apparent that until we are able to answer questions such as these the appeal to balance is not operationally feasible. Of course, the fact that answers to some of these questions involve explicit or implicit value judgements which the economist must take as given does not preclude attempts to find out the nature of public preference patterns.

In addition to the notion of spatial balance, American growth centre policy has been based on the alleged ability of growth centres to help spread economic activity into their hinterlands. This emphasis is current in scholarly publications as well as in government policy. The following passage from a recent article on growth poles is instructive:

The following definition, which I shall adopt for the purposes of this paper, is one which was once agreed upon at a seminar in the regional science department, University of Pennsylvania. 'A growth pole is an urban centre of economic activity which can achieve self-sustaining growth to the point that growth is diffused outward into the pole region and eventually beyond into the less developed region of the nation.' (Nichols, 1969, p. 193).

A related preoccupation in scholarly approaches to growth centre policy in the United States is central place theory and the analysis of urban hierarchies (See, for example, Berry, 1969, and us, 1966). The relationship between this type of study and studies which view growth centres as primarily generators of spread effects is close, because communication of growth involves transmission of stimulating effects through an urban hierarchy. Thus, although growth pole theory originated in France in part as a reaction to central place theory, it is instructive to give somewhat more detailed attention to the incorporation of central place theory into some versions of growth pole theory.

The boundary of a polarized region is often defined as the line beyond which flows and connections are predominantly in some other direction, toward some other pole. A polarized region can exist at any scale and smaller polarized regions will tend to 'nest' within larger ones. The idea of a polarized region is therefore compatible with the central place structure of a hierarchy of cities of ascending size and function, with the growth centres normally being the larger city or cities in the region, at whatever scale is being considered.

The 'range of a good' and 'threshold population' are the key concepts in central place theory. The range of a good denotes the zone around the central place from which persons travel to the centre to purchase the good or service offered at the place. In theory the upper limit of this range is the maximum possible sales radius. Beyond this limit the price of the good is too high because the distance results in too high a price or because of the closer proximity of consumers to alternative centres. The lower limit of the range is the radius that encloses the minimum number of consumers necessary to provide a sales volume adequate for the good to be supplied profitably from the central place. This lower limit is the threshold population. The lowest level of centre performs certain functions or provides certain goods that are limited in number and kind by the limited population within usual range of the centre. The centre of the next highest order performs all the functions of the lower order centre plus a group of additional functions. The next higher order of centre will offer all the goods offered by the first two

levels but will be differentiated from the order just beneath it by a group of goods with greater ranges than those possessed by any of the goods of the next smaller centre. In this manner a hierarchy of centres is determined. It should be pointed out that population by itself is not a good measure of centrality. A large specialized-function town may have only a small tributary area and little influence on its surrounding hinterland. Thus, centrality is generally discussed in terms of centralized services, including administration, culture, health, social services, organization of economic and social life, finance, trade, service industries, the labour market, and traffic.

One problem with the study of urban hierarchies is whether differentiated classes actually exist as discrete entities. The number of classes in any given study may be a result of arbitrary decision, whether conscious or unconscious. Moreover, if discrete classes actually exist they may vary among countries or even among regions within a single country. An even greater problem is identifying precisely how growth is transmitted through a given hierarchy. Central place theory is essentially static, whereas what is required is an explanation of the dynamics of growth in its spatial dimensions. Research into the spatial diffusion of innovations through the urban hierarchy has attempted to remedy this difficulty. It has indicated that new innovations will spread where there is a developed communications network, but that there may be a considerable time lag in the process of diffusion from the innovating centre to its hinterland and to major centres in peripheral areas. However, Richardson accurately maintains that:

'this theory is an inductive theory, developed from particular cases that are not representative of the many different types of innovation. In particular, case studies have been based on new types of consumer goods rather than on industrial innovations and new production techniques. Yet when we are considering the impact of the spatial transmission of technical progress on growth in fringe areas, it is these latter innovation categories that we have in mind'. (1969, p. 316).

For innovations to be adopted in peripheral regions more is required than a well-developed communications system and outward-looking major cities. The peripheral region must have people and organizations capable of evaluating the potential of new innovations and of taking risks. However, the agglomeration economies of the innovating centre will probably keep growth industries and their personnel closely tied to the centre. Branch plants of national firms may be sources of receptivity to new innovations, but they are not likely to flourish in industries with sizeable economies of scale. Also, 'If economic conditions favour

a fast growth in factor inputs at core areas, these same conditions are likely to result in these core areas being the main innovation centres in the economy and to militate against a rapid, wide diffusion of technical progress to outlying regions.' (Richardson, 1969, p. 316).

Finally, a distinction must be made between growth centres and new towns, since the two notions are sometimes associated or made identical. New town proposals call for completely new urban centres whereas growth centre strategies are based on existing places. New town advocates usually espouse self-contained labour markets, whereas the growth centre approach taken in this study stresses the role of job creation for migrants from other – though not necessarily distant – areas. New towns are also generally proposed as means for channeling population growth away from large cities; although growth centres may also perform this function their primary function is to provide economic opportunities for persons from areas with limited opportunities (Alonso, 1970, pp. 37-56). Existing examples of new towns – most notably Columbia and Reston, near Washington, DC – may prove to be successful experiments in town planning but they are primarily vehicles for relocating people within a metropolitan area. Their costs are such that they have little relevance to people in the income categories of migrants from lagging regions.

In sum, then, the approach of this paper assumes not only that economic growth is spatially unbalanced as a matter of fact, but also that growth centre policy should be based on unbalanced growth, favouring urban places which are intermediate to congested areas on the one hand and lagging areas on the other. It also concentrates on growth centres as poles of attraction for commuters and migrants and gives less emphasis than most approaches to growth centres as generators of spread effects. Relatively little use is made of central place theory or new towns concepts.

The following sections deal in turn with the need for human resource investment in lagging regions, problems of congested regions, the nature of intermediate growth centres and migration from lagging areas to growth centres.

## 2. INVESTMENT IN HUMAN RESOURCES

One of the major developments in economics during the 1960s was the immense attention devoted to the significance of investment in human resources, or as some would have it, human capital. Indeed, it now seems almost incredible that at the outset of the decade so little work

had been done in this field. It now is apparent that among the factors which contribute to economic growth the quality of the human input ranks very high. Conversely, one of the principal factors retarding the development of lagging regions is a relative deficiency in human resource development. The lack is particularly evident in areas which have not experienced economic development, as contrasted with declining industrial areas in need of conversion of economic activities and readaptation of the labour force. These points have been developed at considerable length by the author elsewhere (Hansen, 1968; 1970).

The disadvantages which lagging areas encounter on account of deficiencies in the health, education, and training of the labour force are familiar. However, lack of investment in human resources also has adverse effects on the political and business leadership of these areas. Although especially vigorous political and business leadership is needed if improvements are to be made in social and economic problems, 'each of the essential elements in the leadership-technical-expert pattern tend to be relatively weak in distressed areas' (Levin, 1969, p. 203). The political leadership 'is often inbred, weak, and factionalized to the point of near paralysis. A dearth of alternative opportunities combined with decades of selective outmigration have removed young, dedicated, well-educated, and well-motivated men and women whose views extend beyond limited local horizons.' Entrepreneurship in lagging areas 'has been diluted over the years by the dissolution or relocation of stronger local firms that, whatever their faults in "milking" their business and community, nevertheless retained strong local ties and supplied civic direction at critical junctures.' To the extent that new firms are attracted, their managers tend to be persons with brief tenure in branch plants of national firms or else marginal operators dependent on the favour of local politicians. In either case the newcomers have little political impact and they frequently endeavour to retain their ties with other areas of the country. Finally, technicians employed in lagging areas 'are often underpaid, substandard professionals more akin in quality and outlook to local civil servants than to professional staff found in metropolitan communities. The occasional capable elected official finds himself seriously handicapped by the absence of technicians qualified to seek out federal and private outside capital and to design and implement effective programmes.' (Levin, pp. 203-204).

In many cases inadequate investment in human resources has occurred not only because of inadequate local funds but also because available funds have been squandered on attempts to attract industry. Instead of building better schools and using public amenities to attract firms, there has been a pronounced tendency to extend financial inducements



directly to firms and letting the schools wait. What Ralph Widner (1969) has termed 'the era of chasing smokestacks' began with an effusion of brochures representing more or less sophisticated versions of boosterism. However, as communities and states began to outbid one another some 14,000 industrial development organizations came into being to compete for 500 to 750 new plant locations per year. In the scramble to attract industry many communities even went so far as to grant tax moratoria to new firms, thereby losing the principal gain that industrial development was supposed to bring. Thus, instead of upgrading community services, a burden of services to the new plant had to be assumed without any financial benefit from the beneficiary. Moreover, the types of firms that were attracted tended to be labour-intensive (often employing mostly women), slow-growing industries paying low wages. In some cases they would also pull out when another community offered a more favourable subsidy. (See also Sazana, 1970).

Realization of the dangers in trying to subsidize footloose and often marginal firms still has not overcome reluctance to upgrade human resources in many places. Because of the selective nature of outmigration many communities know that better health facilities and schools will only lead to an accelerated exodus of young people. There is something to be said for the people in relatively poor regions not wanting to see the payoff from their investment in mobile human beings go to relatively prosperous areas. Under these conditions poor regions are justified in asking the nation as a whole to support investment in their people, although public works and business-oriented programmes so favoured in regional development legislation are difficult to justify because of better alternatives in regions with greater external economies.

A recent study points up the relative under-investment in human resources in lagging areas (Miller, 1970). Office of Economic Opportunity data on federal investment by county and by standard metropolitan statistical area (SMSA) in 1967 and 1968 are divided into economic overhead capital (primarily public works in the narrow sense) and social overhead capital (investment in human resources).<sup>1</sup> Observations

<sup>1</sup> The economic overhead capital (EOC) and social overhead capital (SOC) breakdowns were based on procedures in Hansen (1965b, pp. 150-162; 1965a, pp. 49-56). In Miller (1970) 62 per cent of the SOC outlays were accounted for by the Department of Health, Education, and Welfare, 11 per cent by the Department of the Interior, and 8 per cent by the Office of Economic Opportunity. Of the EOC outlays, 61 per cent were accounted for by the Department of Transportation, 14 per cent by the Department of Agriculture, and 12 per cent by the Department of Commerce.

were taken for (1) the fifteen largest SMSA's, (2) a random sample of 35 SMSA's with relatively high employment growth, (3) a random sample of 35 SMSA's with relatively low employment growth. These 85 SMSA's comprised 257 counties with a total population of 90 million, about 45 per cent of the national total. Then non-SMSA counties were divided into those qualified for Economic Development Administration assistance and those not qualified. Each of these groups was then divided into relatively high growth and relatively low growth classes. Thus, in addition to SMSA's four sets of 50 counties each were used: (4) high-growth, non-EDA; (5) low-growth, non-EDA; (6) high-growth, EDA; and (7) low-growth EDA. In the cases of both the SMSA's and the non-SMSA counties, 'relatively high growth' and 'relatively low growth' refer to the sample observations and not to all SMSA's or counties in the nation).

The data in Table 1, which shows unweighted average *per capita* federal economic (EOC) and social (SOC) overhead capital outlays, indicate that counties designated for EDA assistance receive less *per capita* SOC investment than any other group except low-growth SMSA's. In contrast they receive much higher *per capita* EOC investment than any other group, a reflection of the public works bias toward lagging areas.

Table 1 *Unweighted average per capita federal SOC and EOC investment, fiscal years 1967 and 1968, by type of area*

| Type of area         | SOC                 |                                 | EOC                 |                                 |
|----------------------|---------------------|---------------------------------|---------------------|---------------------------------|
|                      | Amount<br>(dollars) | SOC as a<br>percentage<br>of OC | Amount<br>(dollars) | EOC as a<br>percentage<br>of OC |
| All states           | 175                 | 57.0                            | 132                 | 43.0                            |
| 85 SMSA's            | 157                 | 63.8                            | 89                  | 36.2                            |
| 15 Largest           | 206                 | 70.5                            | 85                  | 29.5                            |
| High growth          | 137                 | 60.6                            | 88                  | 39.4                            |
| Low growth           | 128                 | 58.7                            | 90                  | 41.3                            |
| 200 Counties         | 172                 | 54.1                            | 146                 | 45.9                            |
| High-growth, Non-EDA | 242                 | 67.0                            | 119                 | 33.0                            |
| Low-growth, Non-EDA  | 178                 | 70.0                            | 114                 | 39.1                            |
| High-growth, EDA     | 135                 | 48.8                            | 142                 | 51.2                            |
| Low-growth, EDA      | 134                 | 39.2                            | 208                 | 60.8                            |

Source: Computed from data compiled by the Office of Economic Opportunity.

In terms of SOC as a proportion of total federal investment, the EDA counties are much lower than any other group, with the low-growth EDA group having by far the lowest value of all, 39.2 per cent.

The *per capita* SOC and EOC values in Table 2 are weighted by population. The per capita SOC value for high-growth EDA counties is below that for SMSA's and for all counties; the comparable value for low-growth EDA counties is somewhat above that for all counties but below that for all SMSA's and each category of SMSA's. However, the SOC-EOC structure of total investments shows that the proportion of SOC is lowest in the EDA counties, and especially in the low-growth EDA counties.

During a visit with a federal official in Washington in the spring of 1970 the author was shown a tabulation by type of area, based on the OEO federal investment data, but more comprehensive in geographic coverage than Miller's tabulations. The official explicitly pointed out that the author's contention that lagging areas are not receiving adequate SOC investment was clearly borne out by the data. However, their use was denied the author by another official on the grounds that 'they have policy implications'.

That the most fundamental problem of lagging areas is under-invest-

Table 2 *Weighted average per capita federal SOC and EOC investment, fiscal years 1967 and 1968, by type of area*

| Type of area         | SOC                 |                                 | EOC                 |                                 |
|----------------------|---------------------|---------------------------------|---------------------|---------------------------------|
|                      | Amount<br>(dollars) | SOC as a<br>percentage<br>of oc | Amount<br>(dollars) | EOC as a<br>percentage<br>of oc |
| All states           | 159                 | 64.1                            | 88                  | 35.9                            |
| 85 SMSA's            | 166                 | 62.8                            | 94                  | 37.2                            |
| 15 Largest           | 181                 | 70.2                            | 77                  | 29.8                            |
| High growth          | 141                 | 56.8                            | 106                 | 43.2                            |
| Low growth           | 156                 | 61.4                            | 98                  | 38.6                            |
| 200 Counties         | 127                 | 55.6                            | 101                 | 44.4                            |
| High-growth, Non-EDA | 102                 | 62.2                            | 61                  | 37.8                            |
| Low-growth, Non-EDA  | 152                 | 63.9                            | 85                  | 36.1                            |
| High-growth, EDA     | 116                 | 53.5                            | 100                 | 46.5                            |
| Low-growth, EDA      | 138                 | 46.9                            | 154                 | 53.1                            |

Source: Computed from data compiled by the Office of Economic Opportunity.

ment in human resources is perhaps indicated best by the return migration which occurs when a new plant does locate in such places. Two clear cases of this phenomenon are seen in Central Appalachia.

In 1956 Kaiser Aluminium and Chemical Corporation built an aluminium reduction and rolling mill at Ravenswood, W. Va., fifty miles north of Charleston. When the plant was first built the company attempted to maximize hiring of workers from the local area. It soon became evident, however, that directly applicable skills and even adequate basic schooling were lacking in the local manpower pool. The kind of worker needed by Kaiser was likely either to be employed, to be a migrant from the area who desired to return, or to be on temporary layoff from some other plant in the Ohio Valley. The direct effects of the plant from the standpoint of lesser qualified persons seeking employment were disappointing. Moreover, the induced employment attributable to the plant's operations could be traced from Ohio (power) to Louisiana (bauxite), but very little could be found in Ravenswood. Only three firms located there to furnish services to the plant. Perhaps 300 jobs were created locally as a result of the town's growth, but they were primarily low-skill jobs in retail trade. New professional jobs were almost all filled by persons from outside the community. Lack of capital and of local entrepreneurship prevented local people from establishing new businesses. In all, about 4,000 new jobs were created in Ravenswood as a consequence of Kaiser's location there. Of this number only about 300 to 500 local persons worked in the plant and another 300 in establishments which came into existence because of Kaiser's growth. In his analysis of the impact of Kaiser on Ravenswood, Irwin Gray (1969, p. 29) concludes that 'More local people could be at work, at the expense of immigrants, if they had had the necessary minimum education or training . . . That more were not hired brings up some pointed questions about education and skills in general.'

Another case is provided by the location of a plant by American Standard, the nation's 90th largest industrial corporation, near Paintsville, Ky. American Standard is the first major firm to locate a manufacturing plant in Eastern Kentucky, but it was clear from the outset that the company counted on migrants who had left for northern cities to return to work in their home region. Among the transplanted Eastern Kentuckians who hoped to find employment with American Standard was Tervert Blackburn, who went to work for Whirlpool Corporation in Ohio and had been gone from Eastern Kentucky for seventeen years. When the Paintsville newspaper ran an item requesting to know if former Eastern Kentuckians would return if work were available, Blackburn circulated it among 60 natives of the region who had

been in Ohio for 15 to 20 years. Blackburn found that 'They all said they would come back if there were work. The people of Ohio don't seem to want much to do with us. They like our work, but that's about it.' American Standard, according to the *Louisville Courier-Journal* 'is gambling that it, too, will like their work, and that the "homing instinct" demonstrated by Blackburn and his friends will provide a loyal and stable work force. Their performance will be critical to Eastern Kentucky's industrial future. "If this doesn't work, no other major industry will go to Eastern Kentucky," said Roger I. McKenzie, who will manage the new American Standard facility.' (Wenz, 1969, p. G 1.)

As of April, 1970, American Standard had received 4,742 applications for employment, with new applications arriving at the rate of about 30 per day. The company expected a total of 7,000 applications by the end of the summer. Although there were initial reports that the plant would employ from 450 to 600 persons, by the summer of 1970, it was only employing 242 persons. Despite the fact that Paintsville is the site of the Mayo State Vocation-Technical School, many of the jobs at American Standard have gone to returned migrants; the plant manager states that 'We've had many move back from Ohio and Michigan. They've learned their skills in the North, and they now can use them here. I've got 500 applications in the file, most of which (sic) would be highly acceptable, highly employable, in a big-city labour market. We're able to be very selective.' (Hawpe, 1970, p. 1.)

In general, then, it is apparent that without substantially more investment in human resources too many of the people of lagging regions will be unprepared not only for employment opportunities in other regions, but even in their own communities should the prospect be opened.

Even though the prospect is not bright, advocates of rural development still claim that every effort should be made to give everyone a job where they now live because migration to big cities results in greater social costs than those that would be involved in implementing their proposals. This raises two questions. First, are big cities really too big? And second, are rural areas and small towns on the one hand and big cities on the other the only alternatives? On the first question it is quite likely that the rural development proponents are correct. However, on the second question they are wrong because they ignore job opportunities in intermediate-sized cities. (Some acknowledge that there may have to be some migration from no longer viable areas to towns, but they are usually quite conservative on the size of the town, as EDA growth centre policy has illustrated.) These propositions will be considered in detail in the following sections.

## 3. THE ROLE OF THE BIG CITIES

Whether or not big cities in the United States are too big cannot be proven. The author has considered both sides of this issue elsewhere (Hansen, 1970, p. 240-248; see also us, 1968) and suggested that they probably are too big in terms of alternatives available to individuals and firms in intermediate-sized cities. More recent evidence indicates that larger cities continue to experience diseconomies of scale in providing public services (Gabler, 1969, p. 425-434), but this argument can never be decisive until it is possible to measure adequately the quality of the services rendered by cities – a prospect which does not seem near at hand. Nevertheless, it is abundantly clear from the media that there is a growing disenchantment with life in our big cities. Thus, Alexander Ganz notes that:

'In recent years it has become fashionable to write off the future of our large cities. Observers point to the shift of population and jobs to the suburbs, the riots of the mid-1960's, the postwar and projected polarization of blacks in the cities and whites in the suburbs, the traffic congestion and pollution, the obsolescence of public facilities and private structures, the poverty of low income families, the fiscal squeeze affecting large city governments, and the deterioration of the environment.' (1969, p. 1).

Since Ganz obviously suffers no myopia about these difficulties it is instructive to consider at some length his defence of the big city. The economies of big cities, he argues, are showing new strength and a larger potential. Since at least 1963 they have not been net losers of jobs. The earlier loss of manufacturing jobs has been more than offset by increased employment in government, business, and personal services. The total number of jobs in eleven large cities for which comparable data were available rose by 5.3 per cent between 1963 and 1967, a rate of growth that exceeded that of the growth of their populations. This phenomenon was experienced by slower growing cities such as Boston, Mass., Philadelphia Pa., and St. Louis, Mo., as well as rapidly growing cities such as Denver, Colo., New Orleans, La., and Washington, D.C. Between 1950 and 1967 the eleven cities lost nearly 600,000 manufacturing and trade jobs, but they gained more than a million jobs in government employment at all levels, finance, communications, advertising, publishing, legal services, arts, theatre, fashions, recreation, medical services, business services, private educational institutions, non-profit institutions, and travel and tourist activities. Productivity has been increasing in all of the cities at a faster rate than in the suburbs.

In general, the industry mix in most large cities now shows as high a proportion of fast-growing activities as the mix of their suburbs. (Ganz, 1969, pp. 8-13.)

Ganz also finds that alleviation of central city traffic congestion is in prospect. Although automobile use and trips are expected to increase substantially in SMSA's, trips to and from the central city may rise only fractionally. Ganz therefore suggests that transportation improvements already underway or planned, including mass transit and highways, may be expected to accommodate the slower growth of travel in large cities. Moreover, he also finds that a notable revitalization of the large cities is underway in response to their new economic functions and to rising personal income and public expenditures. (Ganz, p. 13).

A new beginning of advance and development is occurring as a result of new initiatives in urban ghetto communities (Ganz, pp. 16-19).

'Our studies,' Ganz writes, 'show that despite all of the limitations of life in the ghetto, non-white labour force migrating to the large cities are experiencing notable gains in wages and earnings. The absorption and upgrading of this disadvantaged population is a national problem and a national task which the cities are performing.' (Ganz, p. 25).

The conclusions drawn from this survey are quite explicit. 'The large role of the large cities as producers of goods and services suggests that national growth and welfare would be enhanced by policies and programmes favouring their upgrading and transformation.' (Ganz, p. 15). And again:

'Federal policy should be explicitly designed to favour the large cities and their ghettos through expenditure, grant, loan guarantee and regulatory programmes, in accordance with a measure of their need and their potential contribution to national growth and welfare. Federal policy has recently begun to move in this direction on a number of fronts, but an explicit policy determination would help assure that Federal policy would no longer work at cross purposes.' (Ganz, p. 25).

It is noteworthy that the arguments advanced by Ganz and other advocates of the big city seldom take account of the unquantifiable social costs of urban congestion, nor do they adjust incomes to reflect cost of living. If data limitations preclude our making the relevant subtractions from private pecuniary gain to firms and individuals it is still unwise simply to ignore these problems. If big cities have so many net advantages over other areas, it is curious that a Gallup Poll survey, released in 1968, found that 56 per cent of the respondents would prefer living in rural areas or small towns, if jobs were available. In comparison with a poll taken two years earlier, the proportion of per-

sons expressing a preference for city or suburban living dropped by seven percentage points.<sup>2</sup> The condition of job availability is of course critical, but as Wilbur Thompson (1969, p. 11) has pointed out, 'If blue-collar, middle-income workers should happen to prefer small towns or medium-size cities as places to live and to fish, such a preference is irrelevant as a locational factor. What could be most relevant is that the wives of corporate managers prefer the theater. Under unionism (i.e. equal wages in all places), managers become increasingly free to locate where *they* would like to live.'<sup>3</sup> Thompson and others who emphasize the importance of urban amenities assume that management will always tend to locate in big cities. Up until recently this has been true, but now there is mounting evidence that the managers who determine where the workers will live are increasingly inclined to shun the big city.

*Business Week* recently reported that:

'New York's well-publicized life-style, composed of poor transportation, dirty streets, costly housing, rising crime, strikes, bad air, and bad telephone service, is also getting to business . . . The proverbial Big Apple for many a corporate career has turned out to have a worm in it. Executives are choosing not to take that big promotion that will involve moving to headquarters in Manhattan.' (Feb. 7, 1970, p. 64).

The article goes on to point out that whether they admit it or not, all of the major corporations in New York are having trouble getting people to move there, and that New York offers a preview of what will take place in other big cities if they continue to expand.

Similarly, the *Wall Street Journal* (March 24, 1969, p. 1) reports that business executives are increasingly reluctant to take jobs in big cities because of their expensiveness and discomfort. Executives located in New York, Chicago, and Cleveland are the most inclined to leave for jobs elsewhere, even at lower pay. While there always have been people who cannot endure large cities, 'suddenly, to the growing dismay of corporations, executive talent hunters and management consultants, the metrophobes are legion.'

It may be that the difficulties of the big cities are not so much inherent in their size as in their structure, particularly where it is a question of a bifurcation between blacks in the central city and the white suburbs. Indeed, this position is widely held among the supporters of the big cities. In this event there are two fundamental solutions to the problems of the cities: break down the barriers imposed on the

<sup>2</sup> Cited in a speech by Secretary of Agriculture Orville Freeman before the Conference of Rural-Orientated Industry, Washington, DC, May 13, 1968.

<sup>3</sup> The emphasis is Thompson's.



black population by discrimination and pump more money into the cities to make them more habitable.

But these arguments are not convincing. In the first place nations all over the world are finding that their big cities are too big and an ever-increasing number of urban policies are aimed at checking their growth. While it is obvious that we have structural problems in our cities this should not be an excuse for evading the difficulties of sheer size and density. Moreover, to the extent that we have structural problems they would be easier to deal with if a migration policy would encourage migrants to locate in places other than the big cities. Finally, the argument that big cities can be saved by means of huge doses of federal investment is not in itself appealing; it is the same argument used by the proponents of rural areas and the small towns to save many of them from a natural death, and no doubt Eastern Kentucky, South Texas, and the Indian reservations *could* be made into very attractive places for people and industry if *enough* money were pumped in. The real questions must be posed in terms of spatial opportunity cost: Are there better alternatives in other places? The big cities and the small towns and rural areas obviously need and will receive a great deal of public investment, but it does not seem wise to single them out for special favour; especially when a growth centre strategy based on intermediate-sized cities offers more opportunities in terms of existing external economies than do small towns and rural areas, and fewer diseconomies than do the big cities.

#### 4. INTERMEDIATE-SIZED CITIES AS GROWTH CENTRES

Without speaking of disadvantages, do the big cities have real economic advantages over intermediate-sized cities? The issue here is not one of optimum size but rather of the minimum size required to provide the range of services needed by people and firms and the impact of size on growth potentials.

Brian Berry (1967, p. 18) has found that above a population of 250,000, 'the necessary conditions for self-sustaining growth seem satisfied', and he suggests that the greatest payoff in terms of increasing employment and reducing unemployment would be to use 'the public treasury to enable centres close to this point to achieve self-sustaining growth' rather than to put resources into places much smaller than this maximum.

Wilbur Thompson (1965, p. 24) proposes that there is an urban size ratchet and that when the population growth of an urban area reaches

a critical size of around 250,000 it appears that 'structural characteristics, such as industrial diversification, political power, huge fixed investments, a rich local market, and a steady supply of industrial leadership may almost ensure its continued growth and fully ensure against absolute decline — may, in fact, effect irreversible aggregate growth.'

Since the 250,000 population figure appears so frequently in discussions of efficient city sizes it is instructive to examine what happens to SMSA's as they approach this threshold. Population data at ten-year intervals between 1900 and 1960, as well as estimates for 1965, were used for each of the 231 SMSA's as defined in 1967. Of these, 100 fit into the category of interest in that they had arrived at a population of 250,000 during the 65-year span.

Population was plotted against time for each SMSA. Plots were also made of first differences, percentage changes, and growth on a semi-logarithmic scale. As the cities approached the 200,000 level, 88 of the 100 experienced a growth spurt, that is, a large leap in absolute population. The growth spurts tended to range in magnitude from 50,000 to 200,000 over the relevant ten-year period, with the average being 70,000. This increase of 5,000 to 7,000 per year was substantially larger than any comparable previous growth. The average population at which the spurt began was 173,000 and the average population at the end was 248,000. Some SMSA's continued growing at an ever-increasing rate in subsequent periods while in other cases growth tapered off. Many of the SMSA's reached 250,000 population during their spurt but those that did not arrived at it early in the next ten-year period.

When graphs of the percentage changes were considered, no real conclusions could be drawn. In general, the spurt period occurred during one of the larger percentage changes in growth, but it was by no means always the largest. Percentage change in population during the spurt ranged from highs of 200 per cent to lows of 20 per cent. The average change during the spurt period was between 40 and 50 per cent. The semilogarithmic analysis indicated that city growth was somewhat linear in its natural log form. The spurts also were apparent on a semilog scale. Only twelve cities showed a population decline after the initial spurt period had occurred. (In cities which had not experienced a growth spurt some 26 showed population declines.) These twelve cities could all be classified as being dependent on a single industry or sector for their existence. They did not generally have the 'structural characteristics' which Thompson feels are necessary for self-sustained growth.

The cities which did not undergo a growth spurt showed no similar

characteristics in either their structure or in their growth patterns. The growth patterns ranged from unusually consistent and stable to fairly erratic.

In general, then, the evidence supports the proponents of a growth spurt threshold, but suggests that the threshold tends to be in the 150,000 to 200,000 range, somewhat lower than the usual estimate. However, there is no evidence of automatic self-sustained growth following the spurt.

Australian data indicate that most of the advantages of a city of 500,000 probably also are found in a city of 200,000 but that if a city gets much beyond the 500,000 level external diseconomies are likely to begin to outweigh the concomitant economies. On the basis of Australian experience Neutze (1967, pp. 163, 109-118) suggests that many firms will maximize their profits in centres with populations between 200,000 and one million. In an earlier study Colin Clark (1945, pp. 97-113) examined structural differences in American, Canadian, and Australian cities of different sizes. He concluded that a city of about 200,000 provides practically all important services and that it is 'full grown' with respect to manufacturing at a population level of around 500,000.

Such data as we have with respect to the provision of public services indicate that both small towns and big cities fare worse than intermediate-sized cities. For example, Werner Hirsch (1968, pp. 509-511) estimates that the greatest economies of scale occur in the 50,000 to 100,000 range, whereas the Royal Commission on Local Government in Greater London found the range to be from 100,000 to 250,000. Gordon Cameron (1970, pp. 24-25) finds a 'U-shaped' infrastructure cost curve with the minimum cost lying between somewhat less than 30,000 and somewhat more than 250,000. Critics of such studies usually point out the difficulty of holding the quality of services constant when estimating costs. However, the fact that these studies almost invariably find the range of maximum efficiency to be considerably less than the size of our big cities suggests that until evidence is produced to the contrary the burden of proof lies with the defenders of the big city.

Finally, it is pertinent to note the conclusion drawn by participants at a conference – sponsored by the International Economic Association – in response to the question, 'How large must a successful growth point be?' E. A. G. Robinson reports:

'the general sense of our discussions was that the minimum size of growth points that experience had shown to be successful was nearer to a population of 100,000 than to one of 10,000 and that even 100,000

was more likely to be an underestimate than an overestimate. It must be large enough to provide efficiently the main services of education, medical facilities, banking, shopping facilities . . . Above all, it must be large enough both to provide an efficient infrastructure of public utility services, and to permit the early and progressive growth of external economies for its local industries.' (1969, p. xvi).

In other words, though it is agreed that small towns rarely make viable growth centres, the intermediate-sized city often does have the necessary conditions.

In general, the foregoing material indicates that encouraging (or at least not discouraging) migration from lagging areas may be coupled with a growth centre policy based on external economies in cities in the 200,000 to 750,000 population range. Of course, these are rough indicators, not magic numbers, and the limits could be made more flexible to accommodate cities in, say, the 50,000 to one million range. There is evidence for believing that self-sustained growth is more easy to maintain in a city of 200,000 than in smaller places. On the other hand, external diseconomies may make expansion of alternative locations desirable from an opportunity cost viewpoint after a city passes the 750,000 mark. It has been proposed that the solution to finding an optimum city size consists in finding the 'point at which the economies of scale (or agglomeration) are equalled or exceeded by the diseconomies.' (Reynolds, 1966, p. 21.) Although measurement of these variables is not a realistic prospect for the foreseeable future, the formulation of the problem in this manner is not quite correct. Even if expansion of a big city yielded a positive net social product (economies greater than diseconomies) it would be preferable to have the expansion take place in an intermediate-sized city if the net social product were even greater there. The case for the intermediate city is based on considerable evidence that it has most of the external economies of a big city but that it has not yet become a generator of significant external diseconomies.<sup>4</sup>

<sup>4</sup> Even with respect to amenities one must be careful not to over-estimate the advantages of the big city. New York may offer 300 plays, concerts and recitals in a given week, while a city of 600,000 may offer only 25. Though the overall quality may be better in New York the average person still has time only to take in a fraction of the offerings in the intermediate-sized city. Though there is a wider range of choice in New York, it would be difficult to argue that the cultural advantages of living there are twelve times greater than in the intermediate-sized city. Modern home entertainment equipment also has served to lessen the importance of living in a big city.

## 5. IMPLEMENTING A GROWTH CENTRE STRATEGY

What measures should be taken to implement a growth centre strategy along the lines that have been discussed? Since development policy will focus on areas which are already economically healthy and growing, rather than on areas which have relatively poor growth prospects? There should be more emphasis on measures that will appeal to growing industries and less emphasis on subsidies whose principal appeal is to small firms in slow-growing, low-wage industries. More effort might be devoted to equipping relatively sophisticated industrial sites and less to building water and sewer lines (which may be sorely needed in rural areas, but which should not be a primary concern of an agency whose purpose is to stimulate or accelerate growth). The kinds of tools should be more varied and flexible than those presently applied in small towns and rural areas. The latter need so many improvements in order to make them relatively attractive to firms, especially the bigger and more rapidly-growing ones, that whatever a development agency can do within the constraints of its limited resources is not likely to change greatly the total 'package' of factors that a firm considers when making a location decision. On the other hand, the growth centres that have been proposed here have a large variety of external economies. This means in the first place that a given type of aid extended by an economic development agency would not be so visible as it would be in a lagging area. However, if used wisely, a given type of aid could produce more employment opportunities in the growth centre because it could be combined with these external economies. The development agency should seek out the bottlenecks that are hindering or preventing a firm locating or expanding in a growth centre and attempt to provide the assistance needed to overcome the resistance. The situation may call for a certain type of investment in amenities or in more directly productive infrastructure, or for a labour training subsidy, or for some combination of aid devices. Efforts also should be made to enlist the co-operation of prominent business leaders as is now being done for job creation programmes in the ghettos. A properly dramatized programme of public investment and plant location in growth centres could be a strong force in attracting new economic activity and migrants from lagging areas. Investigations should be made of the degree to which and conditions under which the private sector could be brought into active co-operation in developing growth centres. Large national firms should be involved as well as local business and civic leaders. In any case, it is essential that the aid be made conditional on the extension of job opportunities to persons from lagging regions (and in part to the unemployed and under-

employed residents of the centre).

The emphasis that is given here to the development of intermediate cities as the principal focus for a national regional policy is based not only on the job growth potential of these cities, but also on the fact that problems related to their growth are still amenable to solution. The massive renewal needs of large metropolitan areas can still be avoided by careful planning in growth centres. 'A city of "optimal size"' writes Benjamin Higgins (1968, p. 468) 'must be big enough to be urbane in its range of activities and small enough to provide effective proximity to these activities for its residents, with the available techniques of city planning and transportation.' Unless the government knows what places are going to grow it can provide public facilities only after the demand has appeared. If there is planned growth of relatively few centres, then they can be provided with an integrated and coherent system of public facilities in advance of the demand. Of course, it is not necessary that a growth centre be limited to one city. A system of cities or towns linked by adequate transportation and communications might serve as well or better. Such a system could take the form of a cluster of urban centres or a development axis.

The selective nature of outmigration from lagging areas means that they tend to lose their most vital people – the best workers, the young, the better educated. Thus, outmigration may cause cumulative difficulties in a lagging region, and the benefits from an increase in local employment opportunities may help return migrants more than the local residents. Of course, the positive multiplier effects of any new activity will indirectly benefit the community as a whole, especially if leakages to other areas are minimal. Whatever may be the consequences of outmigration from lagging areas, it is still clear that policies that merely try to check migration – even by attempting to subsidize the industrialization of rural areas – do little service to either the nation or the individuals concerned, at least from an opportunity cost viewpoint. Return migration in particular shows that the real problem of lagging regions is under-investment in their human resources, rather than migration as such, which is a symptom rather than a cause. Hopefully, a national regional policy would aid areas with problems occasioned by outmigration to attain new equilibria with a minimum of friction. The nation may also deem it desirable to aid persons in these areas whose prospects for either local employment or for retraining and migration are not bright; older workers in particular would fall into this category. However, it must be recognized that we are talking here about welfare and not about economic development policy. In any case, public policy for lagging regions should still emphasize active manpower and human

resource programmes, including job information and relocation assistance.

If growth centre policy is to be co-ordinated with human resource development in lagging regions, we need much more knowledge concerning the trade-offs between 'regional development' and worker relocation. As has been argued, many problems of regional development might be dealt with more effectively if they were treated as problems of human resource development and manpower mobility. There is a pronounced need to integrate the research of place-oriented regional economists and the research of labour economists and others concerned with manpower programmes in the broadest sense. For example, regional research could benefit from a number of social science disciplines with respect to our knowledge of attitudes and preferences. When we speak of making spatial resource allocation more rational it is implied that we are in some sense attempting to increase the aggregate level of welfare. But this in turn implies that we know more than we do about the preferences we presumably are trying to satisfy.

#### REFERENCES

- Alonso, W. (1970), 'What are New Towns for?', *Urban Stud.* Vol. 7, No. 1, Febr. 1970.
- Berry, Brian J. L. (1967), *Spatial Organization and Levels of Welfare. Degree of Metropolitan Labor Market Participation as a Variable in Economic Development*. Paper presented to the Economic Development Administration Research Conference, Washington, D.C., Oct. 9-13, 1967.
- Berry, Brian J. L. (ed.) (1969), *Growth Centers and Their Potentials in the Upper Great Lakes Region*. Washington, D.C., Upper Great Lakes Commission.
- Business Week* (1970). Febr. 7, 'For executives, Fun City can be a hardship'.
- Cameron, G. (1970), 'Growth areas, growth centres, and regional conversion', *Scott. J. Polit. Econ.* Vol. 17, No. 1, Febr. 1970.
- Clark, C. (1945), 'The economic functions of a city in relation to its size', *Econometrica* Vol. 13, No. 2, April 1945.
- Gabler, L. R. (1969), 'Economies and diseconomies of scale in urban public sectors', *Land Econ.* Vol. 45, No. 4, Nov. 1969.
- Ganz, A. (1969), *Our Large Cities: New Directions and New Approaches. A Summary of Findings*. Cambridge, Mass., MIT Laboratory for Environmental Studies.
- Gray, I. (1969), 'Employment effect of a new industry in a rural area', *Monthly Labor R.* Vol. 92, No. 6, June 1969.
- Hansen, N. M. (1965a), 'Municipal investment requirements in a growing agglomeration', *Land Econ.* Vol. 41, No. 1, Febr. 1965.

- Hansen, N. M. (1965b), 'The structure and determinants of local public investment expenditures', *R. Econ. Statist.* Vol. 47, No. 2, May 1965.
- Hansen, N. M. (1968), *French Regional Planning*. Bloomington, Ind., Indiana University Press (chapter 1, 7, 11).
- Hansen, N. M. (1970), *Rural Poverty and the Urban Crisis*. Bloomington, Ind., Indiana University Press.
- Hawpe, D. V. (1970), 'Plant at Paintsville is luring workers back to mountains', *Louisville Courier-Journal* June 28, 1970.
- Higgins, B. (1968), *Economic development*. Revised edition. New York, W. W. Norton.
- Hirsch, W. Z. (1968), 'The supply of urban public services' in H. S. Perloff and L. Wingo, Jr. (eds.), *Issues in Urban Economics*. Baltimore, Md., The Johns Hopkins Press.
- Levin, M. R. (1969), *Community and Regional Planning*. New York, Praeger.
- Miller, D. R. (1970), *Public Investment and Regional Economic Development*. Ph. D. dissertation prepared under the supervision of N. M. Hansen at the University of Kentucky.
- Neutze, G. M. (1967), *Economic Policy and the Size of Cities*. New York, Augustus M. Kelley.
- Nichols, V. (1969), 'Growth poles: an evaluation of their propulsive effect', *Environ. Planning* Vol. 1, No. 2.
- Reynolds, D. J. (1969), *Economics, Town planning, and Traffic*, London, Institute of Economic Affairs.
- Richardson, H. W. (1969), *Regional Economics*. New York, Praeger.
- Robinson, E. A. G. (1969), 'Introduction' in: E. A. G. Robinson (ed.), *Backward Areas in Advanced Countries*. New York, St. Martin's Press.
- Sazama, G. W. (1970), 'State industrial development loans: a general analysis', *Land Econ.* Vol. 46, No. 2 (May 1970), pp. 171-180.
- Thompson, W. R. (1965), *A Preface to Urban Economics*. Baltimore, Md., The Johns Hopkins Press.
- Thompson, W. R. (1969), 'The economic base of urban problems' in Neil W. Chamberlain (ed.), *Contemporary Economic Issues*. Homewood, Ill. Richard D. Irwin, Inc.
- US (1966), Department of Commerce, *The Role of Growth Centres in Regional Economic Development*. Report prepared for the Office of Regional Economic Development. Ames, Ia. Iowa State University Department of Economics, 1966.
- US (1968), Advisory Commission on Intergovernmental Relations, *Urban and Rural America. Policies for Future Growth*. Washington, D.C., 1968.
- Wall Street Journal* (1969), March 24.
- Widner, R. R. (1969), *Economic Development and Social Change*. Paper presented to the Government Relations and Planning Policy Conference, American Institute of Planners, Washington, D.C., Jan. 25, 1969.



---

# 27

## The Territorial Basis of National Transportation Planning

John Friedmann and Barbara Stuckey

### Introduction

After a generation of growing centralization of governmental powers, a process which started in the early days of the New Deal, it is becoming apparent that the system of centralized planning and administration we have evolved is neither as efficient nor as responsive to public needs as we once believed. With an annual budget of more than \$200 billion, the Federal Government's present scale of operations is, indeed, unprecedented. Management of a budget this size requires an information-processing capacity of staggering dimensions. Despite heroic efforts to solve this problem, the Federal Government is in some cases acting rapidly but in relative ignorance of the essential facts, and in others with better information but with exasperating slowness.

Partly to overcome this dilemma, federal policies are often formulated and programs conceived and carried out without adequate recognition of the enormous variety of conditions that exist. In a nation of 200 million citizens even small percentages imply impressive absolute numbers. Thus, a policy designed to meet the requirements of a situation in one part of the country may have neutral or even negative effects for substantial numbers of people elsewhere.

These failures, or partial failures, in policy design and program execution have produced a widespread popular demand for bringing governmental operations "closer" to the people. In various ways, and particularly in central cities, people are demanding to be included as active participants in deciding questions concerning their environment. In this way, they hope to bend governmental programs to their specific and local needs as they perceive them.<sup>1</sup>

Finally, the fiscal crisis in state and local governments alike is putting pressures on Washington to share a part of the general revenue and to evolve what has come to be known as a system of "creative federalism." This new relationship between the federal, state, and local governments is beginning to overcome the old antagonism between the states and the central government "whereby the transfer of authority and resources from one legal body to another was deemed to decrease correspondingly the authority and resources of fellow participants . . ."<sup>2</sup> "Creative federalism," says Robert Wood, "grows on the native strength of local institutions."<sup>3</sup>

Reprinted from *Perspectives on Regional Transportation Planning*, ed. Joseph S. DeSalvo (Lexington, Mass.: D.C. Heath and Company, Lexington Books, 1973), pp. 141-175, with permission of the authors and the Rand Corporation. © 1973 by The Rand Corporation.

Given this rising awareness of the need to restructure the governmental process, it is easy to understand why a major governmental sector such as the Department of Transportation should look to the regionalization of some of its operations, particularly of its planning, research, and development activities. Assuming that regionalization in this sense is desirable, what regions would best serve the purposes of the Department? This, essentially, was the question assigned to us by the sponsors of this study. But, simple and straightforward as it sounds, the question actually conjures up two contradictory images of regional delimitation. According to the first, one would search for some uniform set of criteria in order to divide the nation into a set of regions exhaustive of the national territory. Each section of the country would be included in one but not more than one region. Appropriate transport planning would subsequently be suited to the contours of these regions. According to the second image, however, the search would be for a complex pattern of regions corresponding to different purposes and needs. This pattern would not have to cover all the territory of the nation, nor would it have to be thought of as being one-dimensional. Rather, it could be imagined as a pattern of overlapping and superimposed regions of different size and composition.

The second image, however, corresponds more precisely to a question other than the one originally posed: How might the planning and related activities of the Department of Transportation be organized in space so as to meet the needs for transport in different parts of the country more effectively?

An answer must be approached through an examination of the existing pattern of spatial organization, including the division of powers among the several levels of government, the pattern of regions established for economic development and general administration, the distribution of population and activities, and the distribution of levels of economic and social wellbeing.

The present chapter will be concerned with these and other questions. It is divided into three sections. In the first, we explore the relationship between transport planning and various regional concepts that have been proposed. In the second section we turn to an analysis of the major transport problems in the country and lay the groundwork for the third section, where we consider the question of spatial organization for transport planning and conclude with our principal recommendations.

### **The Regional Concept in Transport Planning**

Transport's principal impact is "on the ground." By moving goods and people from one place to another, transport facilitates economic growth at these places and along connecting routes. Once the basic pattern has been laid down, however, subsequent economic growth is produced by forces that are largely independent of the character of the transport system. Although transport must continue to increase as demand increases, it has little generative power of its own.

Transport can perform a developmental role, however. Facilities may be built ahead of demand in the hope that the supply of a service will create its own demand. Although the conditions for this "supply strategy" are not always appropriate, the developmental use of transport is particularly frequent in attempts to solve the problem of "lagging" regions, such as Appalachia, or to open frontier areas to development, as in the building of the transcontinental railroad during the last century.

Particularly in their developmental role, transport services may also act as an organizer of relationships in space, such as patterns of settlement, the location of economic activities, and social interaction. The historical progression of transport technology in the United States has been accompanied by a set of decisive shifts in the spatial organization of the nation's economy. The early importance of water routes gave rise to concentrated developments along the Atlantic seaboard, which later were followed by urban and economic developments on both ends of heavily traveled canals. The change from water to rail added the radial pattern of metropolitan areas as well as the linear structure of inter-regional development. Subsequent reliance on auto transport led to the spreading of metropolitan development on a heretofore unprecedented scale as well as the creation of a complex network of activities along roads and highways between metropolitan centers.

Finally, the introduction of air transport has added to the potential capacity for growth of large centers such as Houston, Dallas, or Atlanta, with many intermetropolitan air connections. Throughout this entire period, some older centers found themselves unable to compete with the new transport technology, became relatively isolated from the mainstream of economic development, and declined in importance.

Although usually thought of in terms of its connective functions, transport may also be a divisive force. The extent of local opposition to the construction of freeways through populated areas is testimony to the real and perceived ability of a large transport line to divide and destroy a community. On a much larger scale, a development such as the proposed Palmdale Intercontinental Airport in Los Angeles County may become a region-forming development. If built, it would in all probability bring about a major reorientation of social and economic relations throughout the Southern California region.

Given these effects of transport services—facilitating, developmental, and space organizing—nothing would appear more logical than to plan the development of transport systems so that these effects are complementary. This raises the question: What sort of area is most suitable for transport planning?

Regional studies have long considered transport links as essential to the definition of regions. The basic spatial patterns in any developed economy can be expressed as the relations between (1) a central city and its surrounding region, and (2) one city region and another.<sup>4</sup> City and region may accordingly be treated as a unit that is characterized by a tight pattern of functional interdependencies and articulated by a system of transport and communication

facilities. These interdependencies are centered at the urban core of the region. Metropolitan-centered regions were eventually defined according to the boundaries of market areas for the city's major trade and service functions.

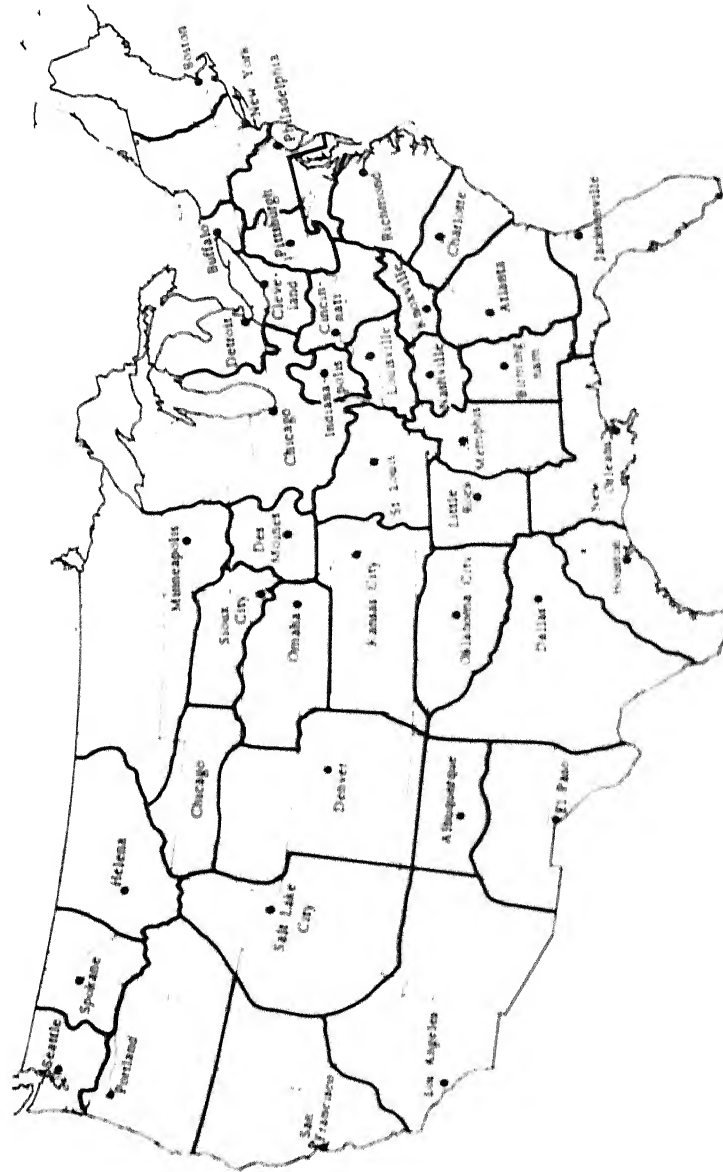
Geographers elaborated this pattern of functional city-region relationships into a theory of "central places."<sup>5</sup> The theory was useful in many ways but deficient in explaining certain types of spatial relations, particularly those arising from manufacturing and certain higher-order services, such as banking and insurance.

Market and service areas often differ substantially in their form and size, so that any given center may be surrounded by as many regions as there are city functions.<sup>6</sup> Although some functional regions may cluster sufficiently to allow one to draw a regional boundary that represents a meaningful average of distance values from the center, other market areas may extend far beyond the immediate localized cluster, even leaping across the contiguous market-service areas of other centers. Moreover, central place theory does not account for the supply areas of urban economic activity, even though the developmental fortunes of two areas may be related "backwards" through the input side as much as "forward" through the pattern of output linkages. In any event, a city may have a close relationship to a region in which it may not be centrally located, if at all.<sup>7</sup>

The question consequently arises of how best to determine the area of influence of any particular urban center in order to construct a set of meaningful urban-centered regions. According to O.D. Duncan, one may take the viewpoint of either the metropolitan center or hinterland. The perspective of the hinterland is reflected in the use of an administrative principle, according to which all land is allocated to the metropolitan area to which it has the closest relationship.<sup>8</sup> Examples of this general procedure are metropolitan regions defined by newspaper circulation (as a surrogate measure for trade area) and Federal Reserve District boundaries (see Figures 3-1 and 3-2).<sup>9</sup> Metro-regions derived in this way take the form of concentric, contiguous areas around the metropolitan center.<sup>10</sup>

If the problem of boundary delimitation is approached from the perspective of the metropolis, however, the relative importance of outlying areas to the functional activities at the center must be considered. Areas nearby may be less important than market or supply areas located large distances away. This approach may therefore lead to sets of discontinuous city-hinterland regions. Moreover, many areas may not be included in any metro-region at all because they are relatively unimportant to metropolitan economies.

The idea of metropolitan-centered regions was officially adopted in 1910, when the Bureau of the Census instituted the concept of the Metropolitan Districts for the purpose of census data collection. A summary of the development and refinement of this concept is contained in the excellent and thorough study by Brian Berry, *Metropolitan Area Definition: A Re-evaluation of Concept*



But the *Manchester Guardian* and the *London Standard* are the only newspapers which are printed by daily newspaper circulation, 1979.

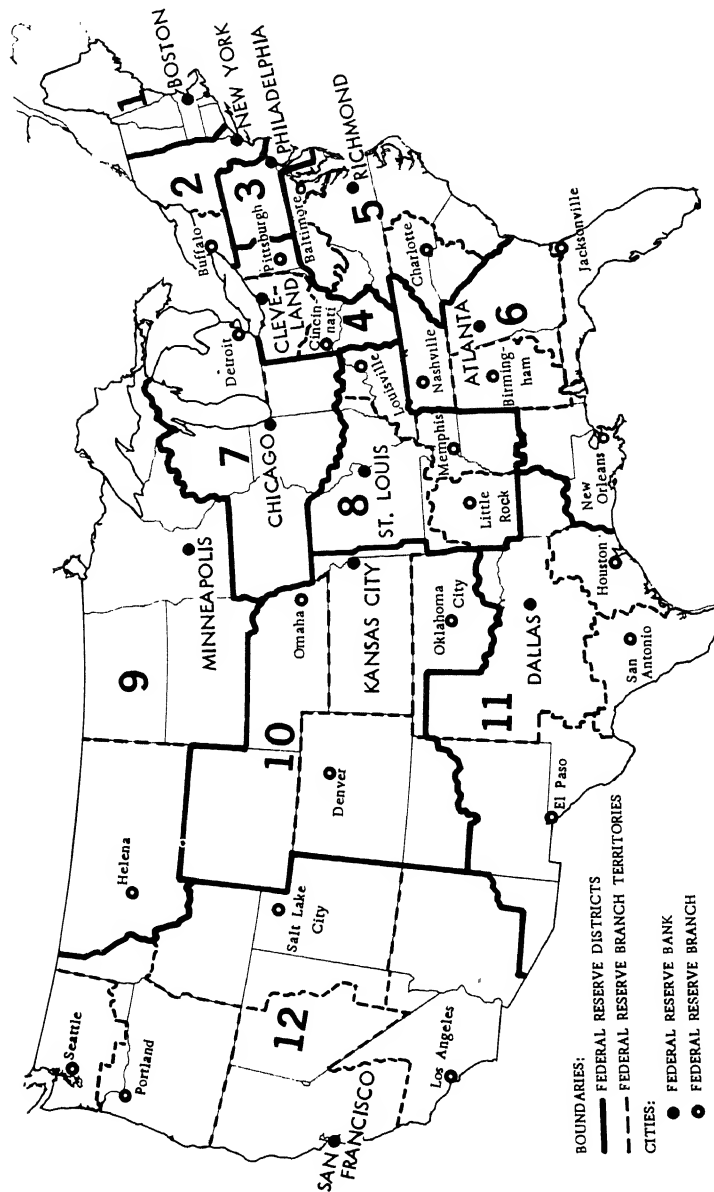


Figure 3-2. Boundaries of Federal Reserve Districts and their branch territories: 1954.

*and Statistical Practice.*<sup>11</sup> The purpose of this report was to "examine the effect of applying alternative criteria to integration of central cities and their outlying areas . . . and . . . to classify the entire United States into a hierarchy of urban, metropolitan, and consolidated areas using criteria of size and of the linkages between places of work, places of residence, and places of shopping."<sup>12</sup> The project was undertaken to define a more useful set of Standard Metropolitan Statistical Areas than was used in the 1960 Census (see Figure 3-3). Berry succeeded in mapping and analyzing commuter data for the entire country; he then applied the concept of a commuter shed to the delineation of metropolitan-centered regions. Following the lead of Karl Fox, he called them Functional Economic Areas. From an economic standpoint, the FEA is equivalent to the labor market of a city. Since the primary data is obtained from commuting behavior, however, it is easy to see how transport facilities will influence a purely economic variable and give to it a spatial dimension. The significance of FEAs for transport planning is discussed extensively in Karl Fox's contribution to this symposium and needs no further comment at this point.

A third type of metropolitan-centered region has been suggested by Constantine Doxiades who refers to what he calls the "Daily Urban System" (DUS).<sup>13</sup> This is based on a radius of approximately 80 miles around urban centers. Such distances, although covered daily by only a few people, are tending to become normal distances for more and more people.

Doxiades argues that establishment of major transport regions for the United States should be based on the boundaries as defined in his mapping of DUS. His criterion would lead to the construction of fewer regions than are derived from the Berry-Fox analysis of FEAs. Moreover, Berry's mapping leaves out less national territory than Doxiades' (see Figures 3-4 and 3-5).

Yet a fourth metropolitan-centered regional concept has been proposed by Friedmann and Miller in an article entitled "The Urban Field"<sup>14</sup> (see Figure 3-6). The concept is based on the idea of metropolitan core areas and what the authors call the "intermetropolitan periphery." Core areas are identified by their relatively high levels of economic and cultural development. Intermetropolitan peripheral areas, in contrast, show relatively low levels of such development. They are, in effect, the problem areas left over in attempts to divide the entire national territory into metropolitan-centered regions; these peripheral areas stand out primarily for their lack of significant relationship to major core regions. Berry's analysis of metropolitan-centered traverses adds empirical evidence to this concept.<sup>15</sup> It shows economic development indices rising to peaks at metropolitan cores (with small dips in the central areas) and falling off rather steeply into the "troughs" of the intermetropolitan periphery as distance from the core increases.

The Urban Field is seen as an evolving spatial pattern that would tend to weld both metropolitan core and the surrounding peripheries into one integrated unit for planning purposes. The Urban Field may therefore be regarded as an

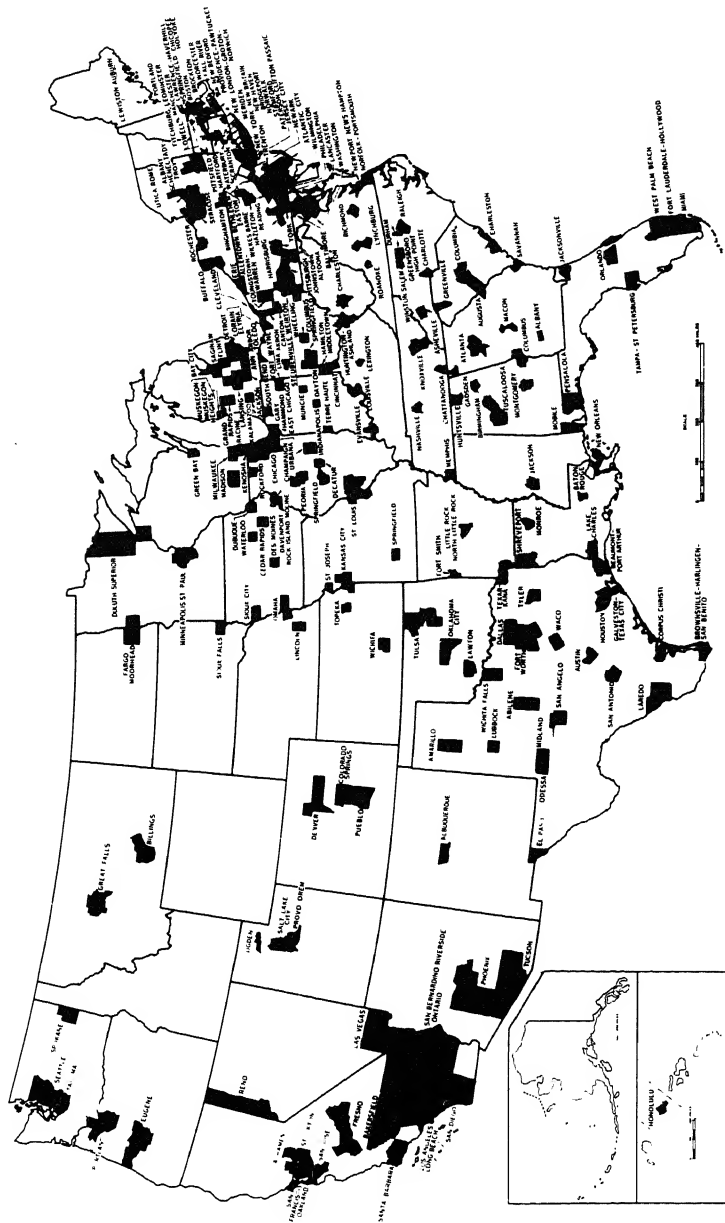


Figure 3-3. Standard Metropolitan Statistical Areas: 1960.



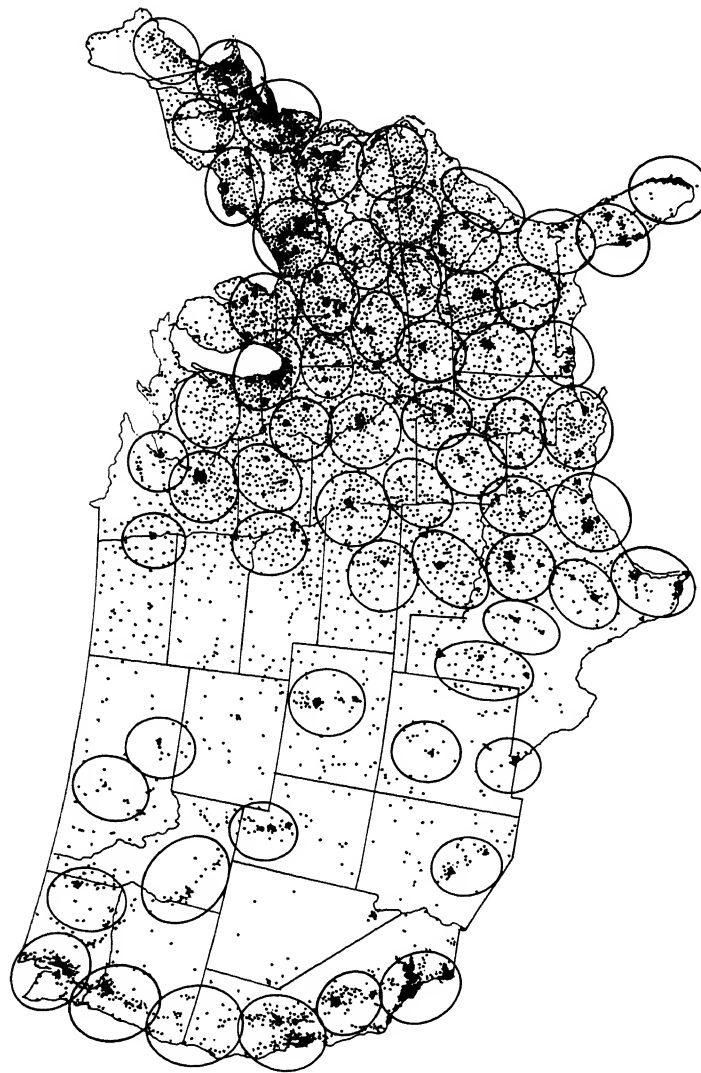


Figure 3-4. Daily Urban Systems.

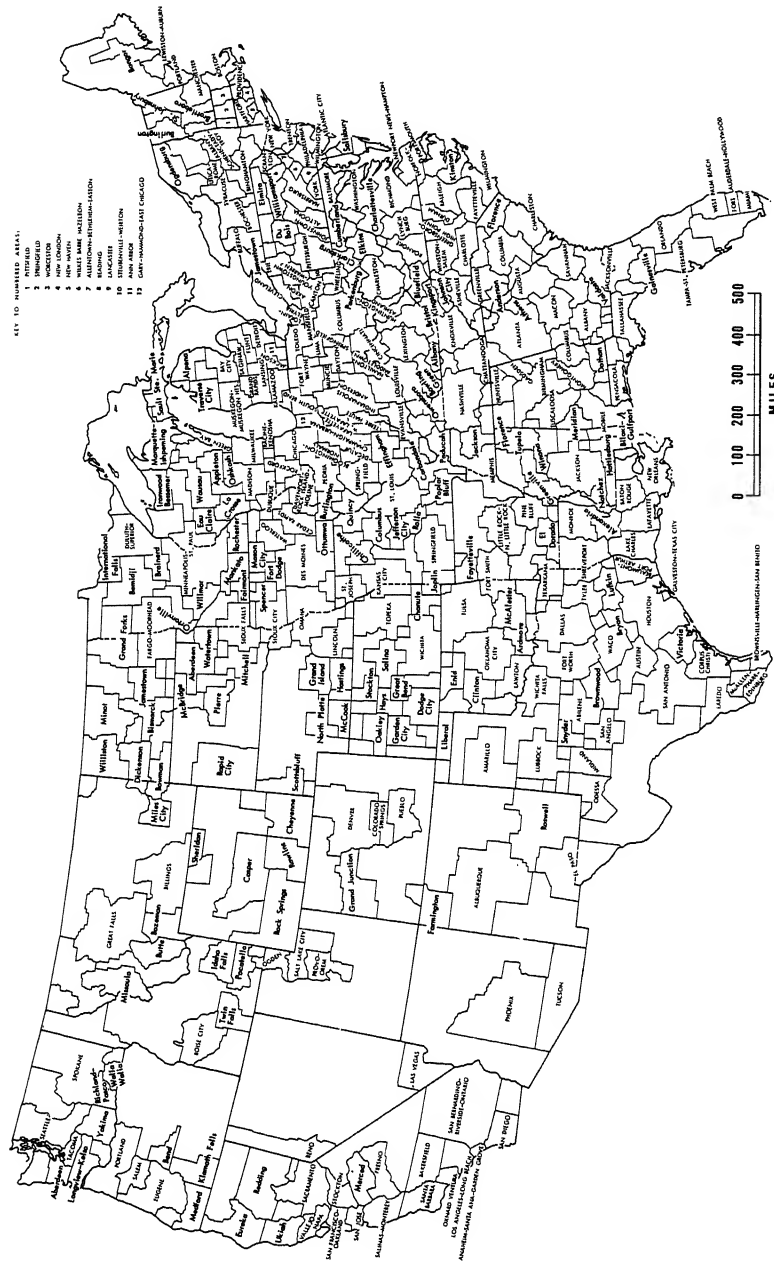


Figure 3-5. Functional Economic Areas.

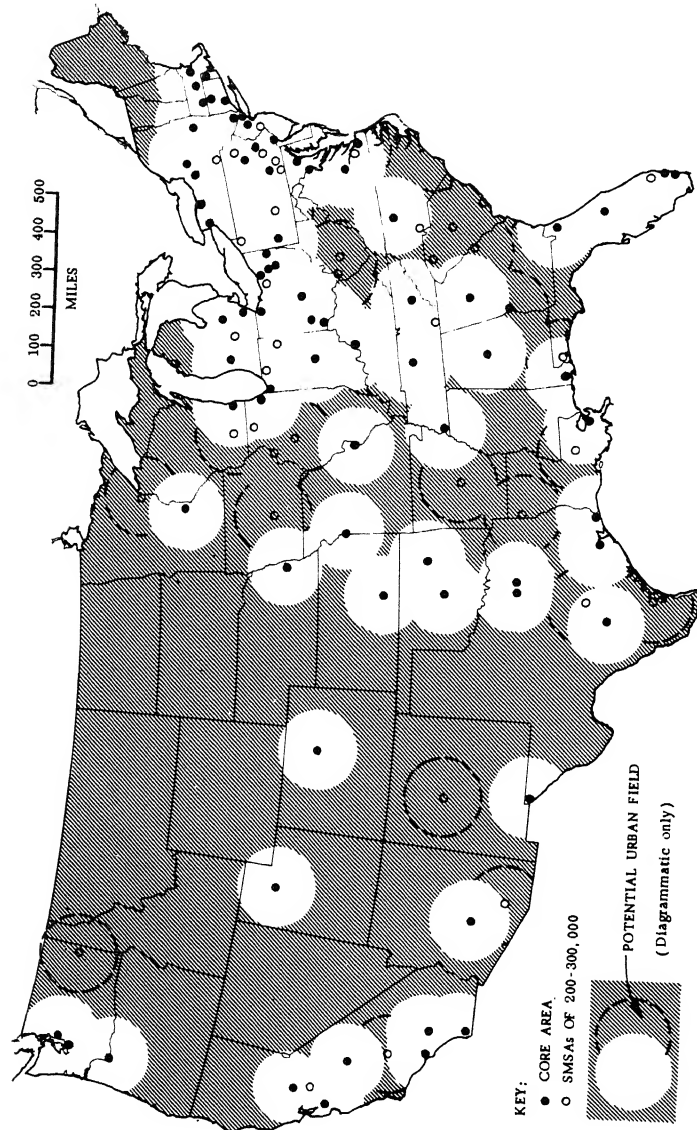


Figure 3-6. Urban Fields.

enlargement of the space for urban living that would extend far beyond the boundaries of FEAs into the open landscape of the periphery. Peripheries would serve as valuable additions of land to densely populated metro-areas, especially for recreation (weekend commuting) and as locations for land-consuming institutions such as schools, hospitals, conference centers, warehouses, and one-story manufacturing units. The authors suggest that Urban Fields be delineated around core areas of at least 300,000 population. Most of the country's population is presently living within the radius of one or another of these fields which occupy approximately one-third of the total land area of the United States.

It is evident from this brief review of regional concepts that the mapping of regional structures will vary with the criterion used. This raises the old question of whether or not regions exist as artifacts or rather as mental constructs that are used to simplify consideration of a complex set of activities. If one could easily delineate regions, one might assert the "real" existence of these regions. On the other hand, no one has ever claimed that regional boundaries can be drawn uniquely and precisely. The relationships that produce a region may exist intuitively or even empirically, but there are great practical difficulties involved in describing the limits of these spatial relationships because, in fact, there are no limits to the spread of interaction. As the authors of a famous public report declared a generation ago:<sup>16</sup>

It seems to be agreed that regional boundaries are usually indefinite, being zones rather than lines. In the majority of instances, therefore, any boundaries which may be drawn will necessarily be arbitrary. It appears, however, that it is desirable to draw boundaries, providing their arbitrary nature be admitted.

Any metropolitan-centered region in the United States is an open region. None of its activities is spatially bounded, even though the intensity of interaction falls off significantly as distance from the central city increases. No matter how carefully regional boundaries may be delimited there is a high degree of overspill into other areas. To trace the boundaries for metropolitan regions, therefore, one must seek the point at which there is the greatest degree of closure for the set of activities considered. Since the choice of activities and their relationships will determine the form of the region, a region is primarily a taxonomical concept.

As we have seen in the development of the Functional Economic Area concept, commuter sheds were the essential element that led to a regional definition. Within each FEA and, to a lesser degree also in the Urban Field, there is a high demand for all forms of ground transport, road and rail, subway and rapid transit. Each of these subsystems must work together in helping to create the spatial pattern of a metropolitan-centered region. Because metropolitan transport systems are tightly integrated and intensively used, there can be little

doubt that within each metropolitan region, however defined, there is need for coordinating transport investments of all kinds. Where the transport system at the metropolitan level works poorly, it will hold back economic growth and make the region a less desirable place for living. The particular character of the transport system, influencing urban form and the structure of communities, will inevitably influence the quality of life throughout the region.

Moving from metropolitan-centered regions to the national level, attention must shift to the relationships among these regions. Although transport assumes a major role in defining these relationships, there is little work that adequately describes the nature of interregional interaction.<sup>17</sup> Both Brian Berry and Otis Duncan have done some research in this area, but their work is of limited relevance to our problem. Berry sets forth a method of analyzing cities as elements within a system of cities.<sup>18</sup> His analytical framework is particularly suitable for studying the intraregional relationships of communities within an Urban Field. Built on the precepts of central place theory, it is not especially useful for describing interactions between two metropolitan-centered regions such as New York and Chicago.

Duncan's work, more empirical in nature, focuses on interregional monetary exchanges.<sup>19</sup> Using flow of funds data, easily collected from the Federal Reserve accounting system and therefore current, Duncan analyzes the role of metropolitan centers in articulating the pattern of interregional relationships. But this work, interesting in its own terms, has not been extrapolated to define the nature of interregional contacts. Although the model Duncan developed suggests that interregional interactions could be identified on the basis of a rather simple set of criteria—distance, population, and regional income—this analysis still remains to be done.

In the absence of a formal model, the pattern of interregional relationships may be described in terms of yet another regional concept, the Urban Corridor. The Corridor occupies an intermediate place in the system of spatial relationships important in the planning of transport systems. It may be considered as a linear extension of metropolitan regions. Normally, it is a large, densely populated area containing several urban nuclei. At the same time, it is an expression of intense intermetropolitan interaction.

An Urban Corridor may be defined as a linear system of urban places that, together with the transit facilities connecting them, constitutes a subsystem of the national space economy.<sup>20</sup> Corridor-centered development is the cumulative result of prior transport development.<sup>21</sup> Urban Corridors are, therefore, clear instances of the region-forming character of transportation. The linear nature of early transport developments—along sea coast, lake shore, and valley routes—is reflected in the heavy concentration of urban developments (and intensive agriculture) in such Corridors as Boston-Washington, Milwaukee-Chicago-Detroit-Pittsburgh, Seattle-Portland, and San Francisco-San Diego. Most of America's urban population lives within the confines of these linear regions.

Since the Urban Corridor is composed of an especially intricate complex of interrelationships, it is a region that requires special solutions to its problems. The transport component of a corridor network of urban activity is, moreover, of special significance, as has been recognized in the Northeast Corridor's approach to intermetropolitan transit.

Nonmetropolitan regions, finally, pose a special problem. In the American context, they are areas that have been by-passed by economic progress. Large in physical area, they are low in population density. They are also poor and subject to large-scale emigration to metropolitan centers and Urban Corridors. As part of the intermetropolitan periphery, they could usefully be considered within the context of the Urban Field. But public policy has preferred to treat them as residual areas, and the Economic Development Administration (EDA) is currently presiding over several Regional Commissions that, established within the major areas of economic depression, are charged with promoting their development. At best, this mission will cushion the shock of continued economic decline.<sup>22</sup> But as long as planning for these regions remains contained within the boundaries of the intermetropolitan periphery—in the Ozarks, the Upper Great Lakes, the Coastal Plains, Appalachia, (Northern) New England, and the Four Corners (comprising parts of Utah, Colorado, Arizona, and New Mexico)—transport planning should be integrated with the general run of developmental planning for these areas. In most cases, however, as in New England, it would be beneficial to think of their development, including transportation, as linked to that of the great metropolitan complexes located outside their boundaries.

In summary, there are three levels at which transport looms as a crucial issue in development policy and planning: the metropolitan-centered region (SMSA, FEA, DUS, or Urban Field); the Urban Corridor; and the Nation. The components of the transport system at subnational levels become articulated and fully integrated at the national level, where concern must focus specifically, though not exclusively, on problems of linkage among Urban Corridors and individual metropolitan regions. Nonmetropolitan areas are best considered as related in development to external core areas. Only where this cannot be done, should they be regarded as yet a fourth level for regional transport planning.

### **Problems of Transport Planning**

The participants in the planning, development, construction, and operation of the national transport system are many. All levels of government—federal, state, metropolitan, and local—perform distinctive roles. Moreover, special transit authorities may have jurisdiction over large metropolitan areas that encompass more than one local government and more than one state. A number of federal agencies, among them the Interstate Commerce Commission, the Federal Aviation Agency, the Interstate Highway Commission, and the Department of

Transportation, have authority to deal with special transport problems. In addition, there are several Congressional committees that deal extensively with transport. The functions of some of these federal participants are represented by counterparts at the state level.

In addition to governmental agencies, there are private participants: railroads, airlines, trucking companies, bus lines, taxi fleets. Although these privately owned systems are publicly regulated by the Federal Government and by scores of agencies at the state and local level, their decisions are seldom made within a common planning framework.<sup>23</sup> Our transport system is therefore managed by a multitude of public and private groups. Although there is a complex pattern of interaction among them, the level of coordination is low. This has led to the development of a transport system that is less than optimal when judged by the multiplicity of needs that it must serve.

The major problems of the national transport system are succinctly stated in the Congressional bill introduced by Senator Magnuson in February 1970.<sup>24</sup> The purpose of the bill is explicit: to provide for the planning and development of a balanced transport system throughout the United States. The bill seeks to encourage "coordinated planning and development of balanced transportation facilities within and between all regions of the United States." Its provisions rest on the explicit assumption that there are, in fact, major geographic and economic regions in the country that, if properly designated, would constitute meaningful areas for planning. Instructions for the delimitation of these regions, however, are brief. According to the authors of the bill, they shall be "major transportation regions," and there must also be geographic, demographic, and economic relationships among the several areas within each region.

In light of the preceding discussion, these apparently simple criteria for regionalization are vague and ambiguous. A clue is perhaps provided in the word "major." Applied to transport regions, it suggests an area of relatively large magnitude, almost certainly larger than that of any single metropolitan-centered region. No constraint, however, is imposed that the relationships of areas *within* such regions must be stronger than relations *among* the regions.

The bill further provides for the establishment by the states, wholly or partially located within such regions, of appropriate multistate regional commissions. Their principal functions are identified as follows:

1. Evaluate transport plans for the region.
2. Initiate and coordinate the preparation of a long-range overall transportation plan for the region.
3. Give due consideration to other federal, state, and local transport planning in the region.
4. Relate transport development to other planning and development activities and needs of the region.
5. Initiate research and development of intercity systems aimed at immediate improvement in intercity passenger service using existing facilities.

6. Initiate research and development of safe and reliable high speed prototype intercity passenger systems.
7. Cooperate with federal, state, and local agencies in conducting or sponsoring research and development programs.
8. Provide a forum for consideration of transportation problems and solutions for the region.
9. Formulate and recommend interregional compacts and other forms of interstate and interregional cooperation when necessary.

Careful scrutiny of what these functions imply leads to a more precise definition of criteria for establishing such commissions. Function 3, for instance, implies that the regional transport commissions are not to have controlling authority over all forms of transport planning in their areas. The bill also calls for popular forums to consider their regional transport problems and possible solutions to them (Function 8). But as a precondition to this "creative" function, and if it is not to become an empty gesture, certain powers of coordination, decisionmaking, and financing would have to devolve to the regions. Although the bill provides for regional funding, it is of prime importance that the funding level be meaningful in terms of the objectives sought. Provisions are made for regional planning (Functions 1 and 2), but not for regional control over most transport investments, which are to be undertaken, as in the past, by a multitude of governmental agencies. By so separating the planning from the implementing functions, the bill in effect creates an impotent planning organization that has little or no political muscle and lacks the economic power to enforce its policies.

Function 4 illustrates the conflict between single-purpose regions (transport planning) and composite regional organizations where all major aspects of development may be closely coordinated. The bill does recognize the advantage of including transport planning in the business of already established regional commissions in the depressed areas. Unfortunately, these areas are for the most part excluded from the major metropolitan economies.

Perhaps the most attractive feature of the bill is the concept of intermodal planning it is seeking to promote. In Senator Magnuson's introduction of the bill in the Senate, his emphasis was unmistakably and almost exclusively on the advantage of intermodal planning.<sup>25</sup> The following are excerpts from the Senator's introductory statement:

It is obvious that we cannot continue building highways in the innocent belief that more roads will relieve congestion. Already, highways and parking facilities occupy as much as two-thirds of the land in our major cities.

... we have failed to provide adequate transportation for those millions of Americans who do not have access to a car.

We must abandon the notion that ... any single mode of transportation can be relied upon to solve the problems of mobility in a prosperous and populous



society . . . that, left to itself, any single mode can cure our transportation difficulties either within or between cities.

The recent myopic reliance on highway development in all parts of the nation has been partially responsible for the current lack of intermodal planning. The highway industry is now well organized and is combatting, as it has done in the past, attempts at intermodal planning and inclusion of highway planning in comprehensive regional plans.<sup>26</sup> This situation begs for new institutional arrangements to put highway development into a proper perspective. One of the shortcomings of the bill is its failure to deal with this problem in a comprehensive way.

Some of the most vital discussion in the hearings turned on the question of intermodal coordination. The testimony of Robert A. Nelson, Federal Executive Fellow at the Brookings Institute, an institution that has done very excellent and extensive research in the area of transport planning, was particularly to the point. Mr. Nelson testified:<sup>27</sup>

Take the airports. One of the reasons we have noncooperation in airport siting disputes is that in fact they are competing not for passenger service but for freight service. There is a real economic reason not to go ahead and cooperate under these circumstances. The airlines are competing for freight service with the railroads and trucks. The railroads are, of course, in the same position with respect to the trucks and airlines. There is a reason not to go ahead and help make the other modes good. If an airline can connect with a truck line rather than another airline, it wouldn't use the airline. That is why I think some authority must be created to insure cooperation. I don't think the various modes will voluntarily move together.

Where, then, should this coordination take place? Nelson concludes that, since large amounts of intercity transport cross state lines, particularly in the East, transport planning and decisionmaking should occur, not at the state level, but at the level of regional or Federal Government.

Senator Edward Kennedy also challenged the bill's authors to put more emphasis on intermodal planning.<sup>28</sup> He went so far as to suggest a reorganization of Congressional jurisdiction over transport programs in order to achieve the desired degree of coordination in the Legislative as well as the Executive Branches of the Government.

Senator Magnuson makes a strong case for the importance of Urban Corridors and cities that three-fourths of the American people will live in cities along such corridors in the near future. He concludes that "satisfying the total transportation needs within and between corridors will involve coordination of different modes and the development of new technology." But his conclusion that "solving the total transportation equation of our transportation regions cannot be the job of a bureaucracy in Washington; the social, environmental, and topographical and demographical factors to be considered will vary too widely

from one region to the next" is contradictory to his premises. For there are numerous and important similarities in the problems of the nation's major Urban Corridors. Solutions to these problems are therefore likely to contain many similarities as well.

There is no compelling reason, for example, why research and development programs should be regionally devised as if the situation in each region were unique. It might be better to guide research activities and their priorities from the national center, even though certain development activities could be carried on in the regions where they are most appropriate. The complete decentralization of research and development would be in direct conflict with the search for solutions that may benefit several regions at the same time. It would seem that prototype development, in particular, ought to remain the responsibility of national transport planning (albeit with the close participation of local and regional groups) in order to avoid duplication, unfair regional advantage, and to provide rapid and full information to all the regions on the successes and failures of new technologies and experimental systems.

In his plea for intermodal planning, Senator Magnuson has established an excellent case for establishing new institutional arrangements to allow for more intermodal communication and control in planning, funding, and implementation. His position was reinforced in the hearings on the bill and by current problems encountered in controlling highway development. The failure of the bill, however, lies in jumping indiscriminately to a regional solution. Examination of the problems clearly suggests new emphasis in intermodal planning, but not necessarily and under all circumstances at the multi-state regional level.

In December 1970, the DOT announced plans for Railpax (now Amtrak), a new government corporation designed to reestablish a functioning system of intercity passenger service (see Figure 3-7). This move is a direct reflection of transport policies for interregional and intermetropolitan travel, which, most appropriately, originated at the national, not regional, level. Current transport problems in the United States focus primarily on intrametropolitan and interregional movements. By laying down a basic trunk line system for rail transport, and by opting for the competitive survival of this system as being in the national interest, Amtrak also increases the probability of third-level bus and air service to the non-subsidized portions of intercity transport linkages. This is clearly a matter for national decision that promises to result in a more "balanced" relationship between the public and private sectors in the provision of transport services. Issues of this kind—and the Amtrak proposal is unlikely to be the last such issue—are clearly best decided at some central level; shifting the onus of decision to regional transport authorities would result in less than optimal solutions for the country as a whole.

### **Spatial Organization for Transport Planning**

The regional approach to development planning has an honorable pedigree. In the United States, it may be traced back to the 1930s. During this early period,

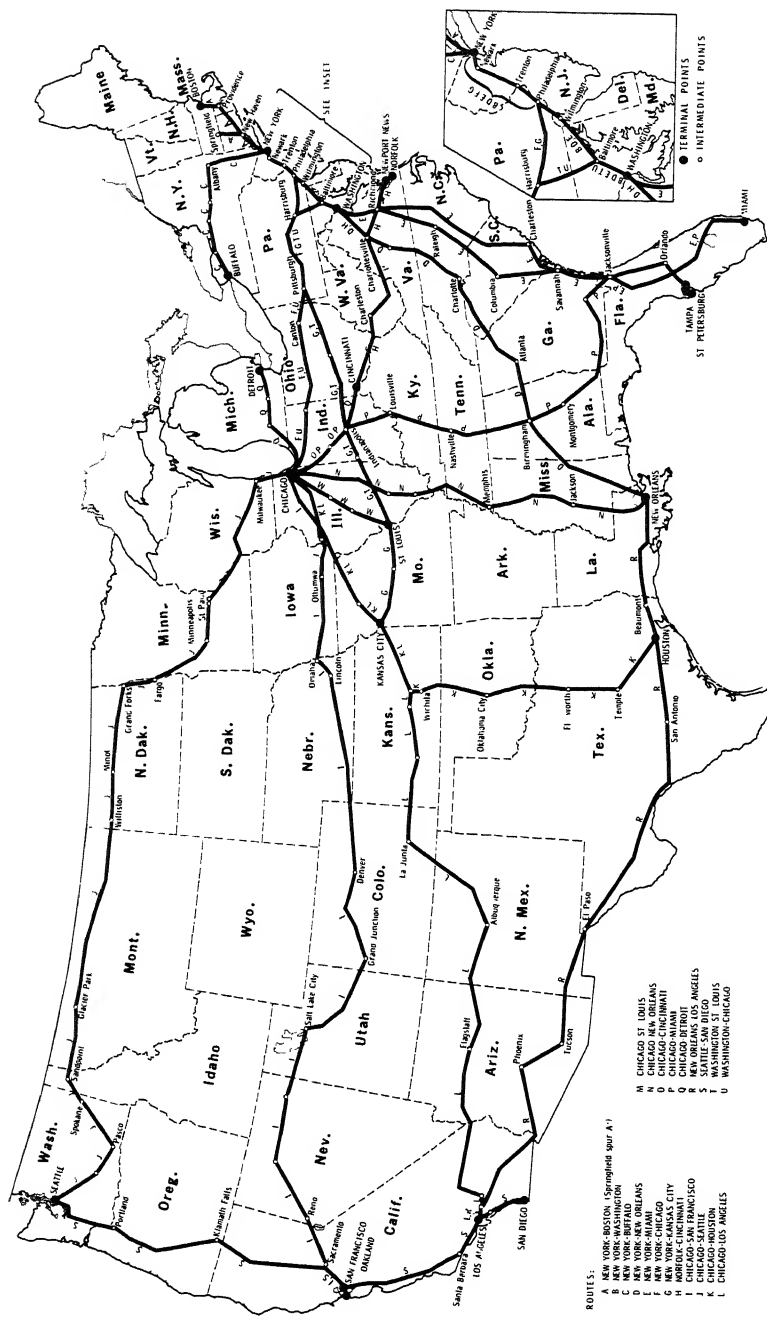


Figure 3-7. AMTRAK Routes.

the river basin was considered the basic unit for planning, and regional policies were aimed at bringing into higher use the water and related land resources of these relatively self-contained physical areas.<sup>29</sup> The most famous example of integrated regional development in this sense was, and still is, the Tennessee Valley Authority. Similar projects, most of them based on the use of water for irrigation, were subsequently undertaken in areas such as the Columbia River Basin. In the light of what has happened since, it is interesting to observe that these early efforts stressed not so much the development of regional economies—which, at the same time, was still a relatively obscure concept—as the “development” of the physical resources and infrastructure of a region. The TVA delighted in the publication of statistics showing the increase in river navigation, the generation of kilowatt hours, the acreage devoted to soil-conserving forage crops, and the number of lake-bound tourists. Integration was to occur in the use of resources, not among the sectors of a regional economy. The Authority did not consider whether its investments would also substantially raise the economic standard of living of the population in the area. This emphasis came only later, after World War II.

Out of this period with its limited vision, however, came a fairly sophisticated and thorough study of regionalization which may still today be read with profit. In 1935, the National Resources Committee presented a report on *Regional Factors in National Planning*. Broad in scope, it includes discussion of the political and organizational implications of regionalization. As its preface announces, “the report deals with important problems of planning and development which overlap state lines or which require the use of combined federal and state powers.”<sup>30</sup>

The Committee made an effective case for regionalization in the United States, where existing state boundaries in most cases have little relevance to the problems of area economies. Further, it established a set of principles to guide its search for a solution to the problem of regionalization.<sup>31</sup> These principles retain their validity:

1. There should be a comprehensive view of the policies of all governments in any given area in order to facilitate solutions to problems of public concern.
2. There is a special responsibility of the Federal Government to ensure coordination of its own agencies dealing with any problem of concern to several governments in any area.
3. There must a sharing of powers among all levels of government.
4. There is a need to decentralize planning of development to a subnational level at which there can be a total review of the problems of any given area by all the agencies concerned.
5. The selection of the area and its planning center should be made with reference to the general coincidence of major planning problems.

No great upsurge of regionalism followed publication of the report. World War II so threatened the survival of the nation, that regional problems appeared of small consequence. Vast regional shifts in labor and industry occurred and brought entire new regions, such as Southern California, into the economy of the nation. This created problems for the older regions, many of which found it difficult to adjust to the changing situation and provided much of the labor supply for the new centers of production. The problem of economically lagging regions was finally recognized during the latter part of the 1950s and eventually led to the creation in 1965 of the Appalachian Regional Commission, and the Economic Development Administration in the Department of Commerce. The EDA was established to coordinate the various federal efforts directed at reviving stagnant regional economies (and in some cases declining urban economies as well) in areas throughout the country. A number of distressed areas were designated as Economic Development Regions. With the exception of New England, however, none contained large, dynamic metro-centers through which a self-sustaining process of regional development might be initiated.<sup>32</sup> (see Figure 3-8).

The most recent step in the evolution of regionalism in the United States had its origin in the growing recognition of the inefficiencies inherent in the centralized administration of bulging federal assistance programs. In a move to simplify administrative procedures, President Nixon directed the Bureau of the Budget to designate a uniform set of regions to be worked with by federal agencies and to establish within each region a single city as headquarters for all the agencies.<sup>33</sup> This has been done. The resulting boundaries are clear reflections of the administrative principle of regionalization (see Figure 3-9). The division was made to allow federal agencies to decentralize some of their functions geographically, to improve physical access of the public to these agencies, and to encourage closer collaboration among them. Initial results are reported to be favorable.<sup>34</sup>

The history of the last 35 years has revealed to us some of the main reasons for which the regionalization of government programs may be undertaken. Starting with the most recent experience, regionalization may be promoted *to achieve improved coordination of programs among different sectors of the Federal Government*. In cases of this sort, the region may be looked upon as a substitute for information.<sup>35</sup> By decentralizing government services and programs, the cost of information exchange between the national agency in Washington and local service areas may be reduced.

Where regionalization is undertaken primarily to achieve better coordination among sectors of government activity, program planning and implementation procedures will tend to predominate; it is shortrun considerations that will count. On the other hand, basic policy decisions, resource allocation among sectors and programs, and expenditure controls will remain in the hands of central bureaucrats.

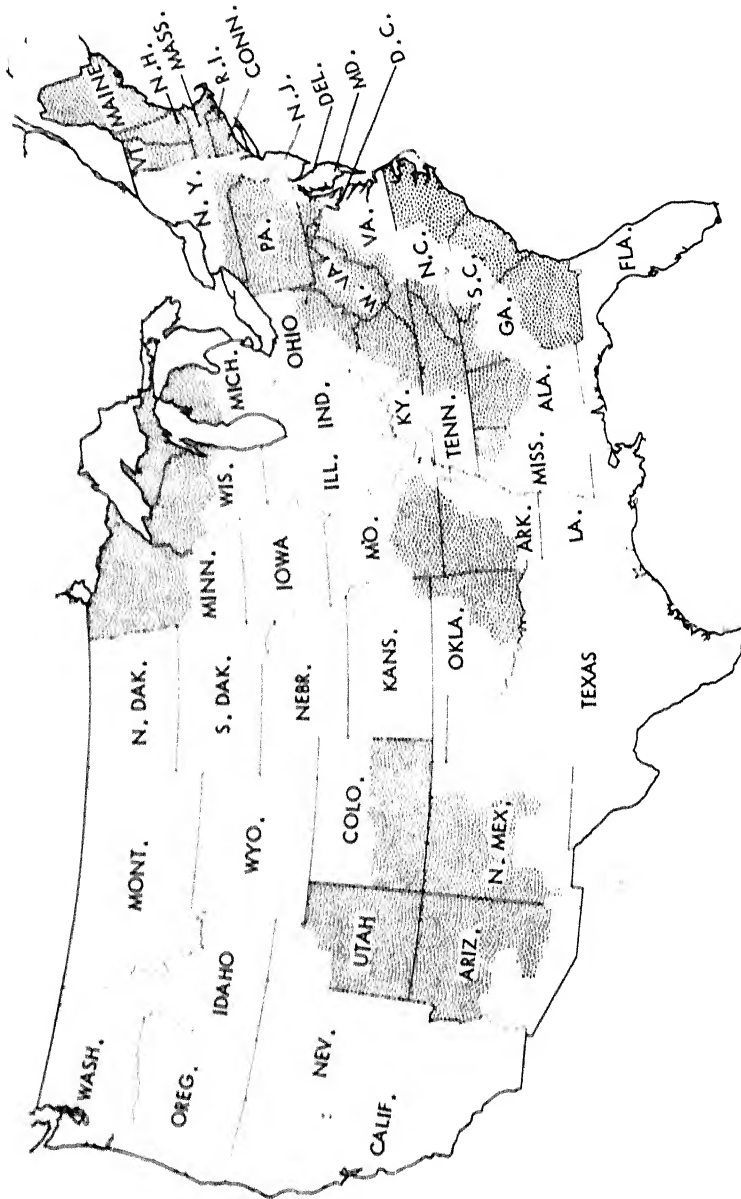


Figure 3-8. Economic Development Regions.

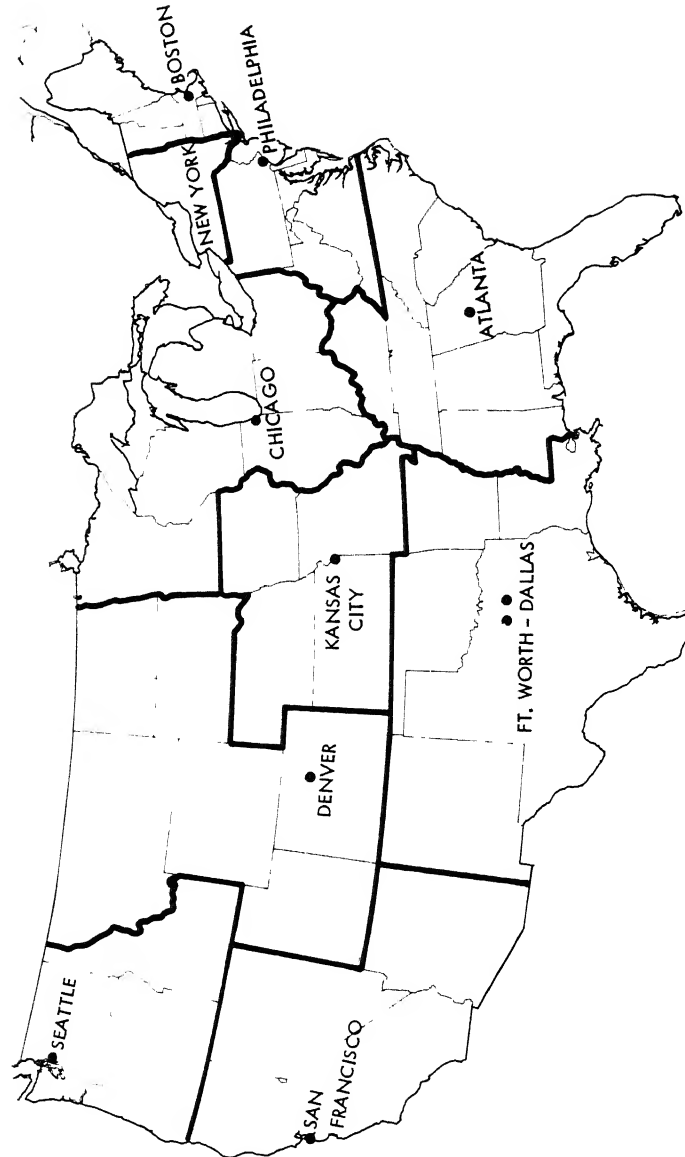


Figure 3-9. Administrative Regional Areas: 1970.

A second reason for regionalization, more ambitious than the first, is *to devolve to regional organizations authority to recommend development policies to the Federal Government*. Such recommendations, which would be evolved in close cooperation with the people of the region through mechanisms such as citizen representation, public hearings, and the like, would have a primarily indicative value for central planners. The Regional Transportation Commissions established under the proposed National Transportation Act (S. 2425, 91st Congress, 1st Session) would be of this sort. As a practical matter, however, such policy-recommending authority is often joined to powers of implementation. A particularly strong policy-program linkage was accomplished with the Tennessee Valley Authority. In this case, authority was devolved to an autonomous agency of the Federal Government, responsible only to the President and to the Congress. This early experiment with the regional administration of development programs was not repeated, however. A looser and politically more responsive arrangement is represented by the five Regional Commissions set up by the Economic Development Administration. In addition to being charged with analyzing the problems of their respective area economies, each Commission is expected to develop an overall strategy for promoting the area's long-term growth and to work with federal, state, and local governments for the implementation of specific programs. To strengthen its bargaining position in the implementation phases of its work, each Commission commands certain investment funds of its own.<sup>36</sup>

Essentially, the devolution of authority over policy matters is yet another attempt—this time at the level of objectives and general instrumentation—to achieve greater coordination among the sectors of government activity. The population living in a region is often thought to have a substantial degree of social cohesion and to be capable of formulating and expressing a mutual interest in the solution of problems specific to their region. Whether this is, in fact, true for every kind of region may be doubted. The traditional broad cultural regions of the United States, such as the Southeast, Northwest, and Northeast, though commonly referred to in everyday conversation, scarcely represent optimal areas for the articulation of a common point of view (see Figure 3-10). This should be borne in mind when considering large multi-state regions for transport planning.

A third reason for undertaking the regionalization of government programs would be *to render such programs more responsive to the wishes of regional populations*. Or, to put it more strongly, it would extend to regional populations a greater measure of political control over the development of their economies. In the extreme case, this would imply some form of regional government. Because of the political structure of the United States, however, the establishment of regional governments across state boundaries has never been attempted. There is, nevertheless, a widespread belief that the regionalization (and decentralization) of government programs somehow brings government “closer” to the



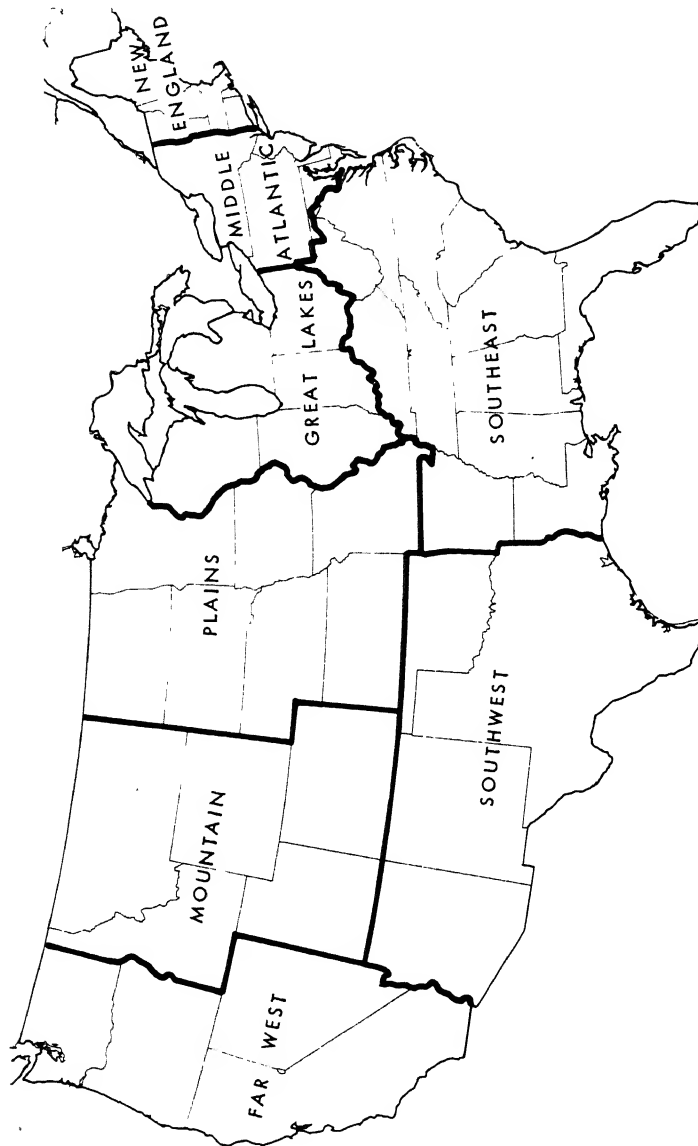


Figure 3-10. Regional grouping of states.

people. As Morton Grodzins has argued in a persuasive essay, this belief may be mistaken.<sup>37</sup> Regional Transportation Commissions, for instance, because of their probable reliance on technical criteria in decisionmaking, their low political profile, the complexity of state-federal negotiations in the context of regionwide bargaining, and their physical location at some central place within a large multi-state region may in fact be less accessible and responsive to the people of the region than Transportation Departments operating within the political framework of a given state. Moreover, even where accountability is built into the structure of such Commissions, it is not clear whether the political will of a regional population group should always prevail in the event of a conflict with national purposes, especially when the funds for implementation come from the Federal Government.

Given this background, the regionalization of transport planning may now be considered in more detail. Earlier in this chapter, we suggested that a useful distinction could be made between the developmental and facilitating functions of transport investments. In case of the former, the extension of transport facilities is regarded as a means for promoting the economic growth of an area and, ultimately, for generating its own demand. The application of this principle is most appropriate in the context of planning for economically depressed regions. The emphasis here would be primarily on relating transport to other kinds of public and private investment in an inter-sectoral or "comprehensive" strategy for regional development. The Economic Development Regions of the country are the logical foci for integrating transport planning with general development planning in this manner. The provision of the proposed National Transportation Act to designate, for purposes of the Act and under certain conditions, an existing Regional Commission as the Regional Transportation Commission is therefore eminently desirable.<sup>38</sup> Its principal shortcoming lies in the fact that nearly all of the depressed areas of the United States are so delimited as to exclude major metropolitan growth centers.

In all other areas of the country, however, transport plays primarily a *facilitating* role in the development process; it is supplied in response to an already existing or projected demand that is generated independently of the provision of transport services. In this case, an inter-sectoral planning approach is needed less than one which emphasizes intermodal coordination.

A further distinction is here required. In many situations, the provision of transport services within a given area does not require repeated and successive decisionmaking. A single project may be planned and actually constructed, but it will be many years before another proposal would make it necessary for this same set of participants to join in a common endeavor. An instance of this would be the joint planning of transport facilities between Southern California and Las Vegas. Cases of this sort can usually be managed quite well through specific *ad hoc* arrangements. They do not require a permanent regional framework for planning and decisionmaking.

In many other situations, however, the complexity of the transport system and its need for adaptability to changing circumstances is such that planning and decisionmaking require continuing attention. This occurs primarily at the level of metropolitan-centered regions, the states, Urban Corridors, and the nation.

In *metropolitan-centered regions*, it would be desirable to integrate transport planning with already existing Metropolitan Associations of Government in order to strengthen coordinating powers at this level. As we have seen, several definitions of metropolitan-centered regions have been proposed, such as the Functional Economic Area, the Urban Daily System, and the Urban Field. Although it is probably true that better planning determinations can be made for larger areas from a strictly technical point of view, administrative convenience may outweigh any advantages of this sort. The jurisdictional area of Metropolitan Government Associations would be a sufficiently good start for achieving greater intermodal coordination in the densely built-up areas of the nation.

For much the same reason, the *states* would appear to be the logical focus for intermodal planning. Less than one-third of the states in the union have to date created departments that might serve as counterparts to the Federal Department of Transportation. Planning and development of each mode is carried on independently by agencies that serve and are supported by special client groups. This arrangement makes it extremely difficult to arrive at a comprehensive transport policy and leads to numerous conflicts that could be avoided or, at least, more readily resolved within the context of a single administrative body. By so bringing the various state transport services together, greater political accountability could be achieved, and tax dollars might be saved on a substantial scale.

A number of *Urban Corridors* fall entirely within the boundaries of a single state (e.g., California, Florida), so that planning for transport facilities could readily be absorbed by the appropriate state agency. In a few instances, however, especially in the Northeast Corridor and in the corridor areas linking Milwaukee with Chicago, Detroit, Buffalo, and Pittsburgh, special transport planning commissions could be established. The transport problems of these corridors are persistent and complex, and intercity problems are particularly critical. Urban corridors could therefore be used to structure and plan for complicated sets of transport problems that are primarily intermodal in character. The provisions of the proposed National Transportation Act are particularly appropriate to corridor planning. In addition to state representatives, members of the Metropolitan Government Associations contained within each corridor could serve on the Regional Transportation Commission to provide for a smoother coordination of all the interests concerned.

Finally, all other problems of inter-metropolitan transport may best be solved at the *national level*. This conclusion should not stand in the way of the decentralization of certain routine administrative functions to the newly established administrative regions of the country. The Bureau of the Budget's recent

division of the nation into major administrative areas is perhaps as good as any. The DOT might be well advised to adopt this division for routine purposes, such as grant processing, simply because the advantages of coordination with other agencies might exceed any disadvantages resulting from a possible suboptimal delimitation for the administration of transport programs. In fact, all of the nation's corridor areas, except the Northeast, are included in their entirety within one of the newly established administrative regions. But the formulation of policies, the allocation of resources, and the coordination of rail, highway, and air transit between metropolitan-centered regions and urban corridors are not primarily problems of a regional nature. The recent creation of Railpax (now Amtrak) is a reflection of the national nature of these problems.

#### Summary of Recommendations

1. The definition of regions for transport planning in the U.S. is best approached through an ordering of specific activities in space.
2. Developmental functions of transport services are best planned in close coordination with other developmental considerations. The Economic Development Regions of the country are logical foci for evolving comprehensive strategies of regional development.
3. The facilitating roles of transport services arise in connection with an already existing or projected demand that is generated independently of the provision of these services. In this case, an intersectoral planning approach may be needed less than one which emphasizes intermodal coordination.
4. Intermodal coordination is called for especially where (a) the provision of transport services within a given area does not require repeated and successive decisionmaking and (b) the complexity of the transport system requires continuing attention. In the first case, planning can generally be accomplished on an *ad hoc* basis, without the designation of special transport regions. In the second case, however, several distinct situations must be considered.
  - at the level of *metropolitan-centered regions*, transport planning should be integrated with already existing Metropolitan Associations of Government in order to strengthen their coordinating powers;
  - at the level of the *states*, improved intermodal coordination may require the creation of Departments of Transportation capable of bringing together the various state transport services, of achieving greater political accountability, and of securing a substantial saving of resources;
  - at the level of interstate *Urban Corridors*, such as the Northeast or the corridor areas linking Milwaukee with Chicago, Detroit, Buffalo, and

Pittsburgh, special transport planning commissions might be established.

5. All other transport planning activities could continue to be carried out centrally at the national level, except certain routine functions, such as grant processing, might be decentralized to the already existing administrative regions for better coordination with other federal programs.

#### Notes

1. See the papers produced for "Can New York City Be Governed? A Symposium on the Realities of Decentralizing New York City Government," held at the House of the Association of the Bar of the City of New York, December 3-4, 1970.
2. "Intergovernmental Relationships in an Urbanizing America," in D.P. Moynihan (ed.), *TOWARD A NATIONAL URBAN POLICY*, Basic Books, New York, 1970, p. 47.
3. *Ibid.*, p. 48.
4. J. Friedmann, "The Concept of a Planning Region," *LAND ECONOMICS*, Vol. 32, 1956, pp. 1-13. Reprinted in Friedmann and Alonso (eds.), *REGIONAL DEVELOPMENT AND PLANNING: A READER*, The MIT Press, Cambridge, Massachusetts, 1964, pp. 497-518.
5. B.J.L. Berry, *GEOGRAPHY OF MARKET CENTERS AND RETAIL DISTRIBUTION*, Prentice Hall, Englewood Cliffs, New Jersey, 1967.
6. O.D. Duncan, et al., *METROPOLIS AND REGION*, The Johns Hopkins Press, Boston, Massachusetts, 1960, p. 39.
7. *Ibid.*, p. 44.
8. *Ibid.*, p. 249.
9. Figure sources listed at end of this chapter.
10. O.D. Duncan, *op. cit.*
11. U.S. Bureau of the Census (Brian Berry, author), *METROPOLITAN AREA DEFINITION: A RE-EVALUATION OF CONCEPT AND STATISTICAL PRACTICE* (rev.), Bureau of the Census, Working Paper No. 28, Washington, D.C., 1969.
12. *Ibid.*, p. iii.
13. *NATIONAL TRANSPORTATION ACT: HEARINGS BEFORE THE COMMITTEE ON COMMERCE*, 91st Cong., Second Sess., S.924 and S.2425, Serial 91-69, U.S. Government Printing Office, Washington, D.C., 1970, p. 195.
14. J. Friedmann and J. Miller, "The Urban Field," *JOURNAL OF THE AMERICAN INSTITUTE OF PLANNERS*, Vol. 31, No. 4, November 1965, pp. 312-320.
15. U.S. Bureau of the Census, *op. cit.*
16. National Resources Committee, *REGIONAL FACTORS IN NATIONAL*

PLANNING, U.S. Government Printing Office, Washington, D.C., 1935, p. 145.

17. J. Friedmann, "The Concept of a Planning Region," *op. cit.*, p. 514.

18. B.J.L. Berry, "Cities as Systems Within Systems of Cities," in Friedmann and Alonso (eds.), *REGIONAL DEVELOPMENT AND PLANNING: A READER*, The MIT Press, Cambridge, Massachusetts, 1964, pp. 116-137.

19. O.D. Duncan, *op. cit.*, p. 133.

20. C.F.J. Whebell, "Corridors: A Theory of Urban Systems," *ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS*, Vol. 59, No. 1, March 1969, pp. 1-26.

21. *Ibid.*

22. N. Hansen, *RURAL POVERTY AND THE URBAN CRISIS: A STRATEGY FOR REGIONAL DEVELOPMENT*, Indiana University Press, Bloomington, Indiana, 1970.

23. *NATIONAL TRANSPORTATION ACT: HEARINGS*, *op. cit.*, p. 236.

24. *NATIONAL TRANSPORTATION ACT OF 1969*, S.2425, 91st Cong., U.S. Government Printing Office, Washington, D.C., 1969.

25. "Announcement of Hearings on S.2425, The National Transportation Act," *CONGRESSIONAL RECORD - Senate*, U.S. Government Printing Office, Washington, D.C., February 20, 1970, pp. S.2127-2129.

26. Witness the recent defeat of Proposition 18 on the November 1970 ballot in California. The measure proposed to divert up to 25 percent of gasoline tax revenue from highway construction to the development of rapid transit systems. Although the proposition had bipartisan support in the legislature, support of the Governor, and widespread citizen support, the members of the highway lobby waged an extensive and expensive campaign based on the false claim that passage of the measure would lead to higher taxes. The role and the extent of monetary outlay by the members of the highway lobby are now being investigated by state officials. See also, T.A. Morehouse, "The 1962 Highway Act: A Study in Artful Interpretation," *JOURNAL OF THE AMERICAN INSTITUTE OF PLANNERS*, Vol. XXXV, No. 3, May 1969, pp. 160-168.

27. *NATIONAL TRANSPORTATION ACT: HEARINGS*, *op. cit.*, pp. 112-113.

28. *Ibid.*, pp. 224-228.

29. Friedmann, "The Concept of a Planning Region," *op. cit.*, p. 498.

30. National Resources Committee, *op. cit.*, preface.

31. *Ibid.*, p. vii.

32. J. Friedmann, "Poor Regions and Poor Nations: Perspectives on the Problems of Appalachia," *SOUTHERN ECONOMIC JOURNAL*, Vol. 32, No. 4, April 1966, pp. 465-473.

33. Bureau of the Budget, *SIMPLIFYING FEDERAL AID TO STATES AND COMMUNITIES*, (Executive Office of the President), Bureau of the Budget, Washington, D.C., March 1970.

34. Ibid.
35. Walter Stöhr, "The Role of Regions for Development in Latin America," *REGIONAL STUDIES*, Vol. 3, pp. 81-90.
36. Hansen, op. cit., p. 109.
37. Morton Grodzins, *THE AMERICAN SYSTEM, A NEW VIEW OF GOVERNMENT IN THE UNITED STATES*, Ch. 7, "Local Is As Local Does," Rand McNally and Co., Chicago, Illinois, 1966.
38. NATIONAL TRANSPORTATION ACT OF 1969, op. cit., Section 5 (b).

### Bibliography

Duncan, O.D., et al., *METROPOLIS AND REGION*, Johns Hopkins Press, Baltimore, 1960. (See especially pp. 26-45 and 90-104.)

Major work done with focus on the role played by metropolitan centers in defining interregional relationships. Also includes some analysis of the economic relationships between metropolitan centers and their surrounding areas.

Dziewonski, Kazimierz, "Economic Regionalization," *GEOGRAPHIA POLONICA*, Vol. 1, 1964, pp. 171-185.

Brief, comprehensive look at regionalization from economic point of view. Includes specific discussion of delimitation of economic regions.

Friedmann, J., "The Concept of a Planning Region." *LAND ECONOMICS*, Vol. 32, 1965, pp. 1-13. Reprinted in Friedmann and Alonso (eds.), *REGIONAL DEVELOPMENT AND PLANNING: A READER*, The MIT Press, Cambridge, Massachusetts, 1964.

An historical summary of the concept of area delimitation for planning purposes in the U.S. Although the focus is on planning activities of a regional rather than city scale, the article concludes with a plea for the integration of city and regional planning within areas that are designated to include both central city and surrounding region.

\_\_\_\_\_, *REGIONAL DEVELOPMENT POLICY*, The MIT Press, Cambridge, Massachusetts, 1966.

Chapter 3 provides an especially direct discussion of the problems of delimitation.

\_\_\_\_\_, and J. Miller, "The Urban Field," *JOURNAL OF THE AMERICAN INSTITUTE OF PLANNERS*, Vol. XXXI, No. 4, November 1965, pp. 312-320.

The urban field is an expansion of the metropolitan region. By including the peripheral areas within the concept of urban core region, a new spatial order is proposed which would unify core and periphery into a single matrix.

Fox, K., "The Functional Economic Area: Delineation and Implications for Economic Analysis and Policy," March 1965 (mimeo).

Development of criteria to delimit metropolitan-centered regions, called Functional Economic Areas.

Gajda, R.T., "Research in Methods of Delimitation of Economic Regions in Areas of Extensively-Dispersed Economy: A Case Study of Northern Canada," *GEOGRAPHIA POLONICA*, Vol. 8, 1965, pp. 27-37.

Specific case study of Northern Canada. Limited possibilities for extrapolation to more common urban centered economies.

Harris, C., "Methods of Research in Economic Regionalization," *GEOGRAPHIA POLONICA*, Vol. 4, 1965.



Very academic work includes extensive bibliographic work on regionalization, 291 entries. Also includes discussion of types of regions.

Macka, M. (ed.), *ECONOMIC REGIONALIZATION*, (Proceedings of the 4th General Meeting of the Commission on Methods of Economic Regionalization of the International Geographical Union, September 7-12, 1965, in Brno, Czechoslovakia): Publishing House of the Czechoslovak Academy of Sciences, Prague, 1967.

Collection of papers by variety of nationalities. Includes section on typology and structure of economic regions. Articles are in French or English.

National Resources Committee, *REGIONAL FACTORS IN NATIONAL PLANNING*, U.S. Government Printing Office, Washington, D.C., 1935.

This report remains the most all-encompassing analysis of the implications of regionalization for the U.S. Although written 35 years ago, the book deals thoroughly with issues that remain frequent topics of debate. Discussion includes spatial, political, and organizational implications of regionalization. See especially Chapter XV.

Stöhr, W., "The Role of Regions for Development in Latin America," *REGIONAL STUDIES*, Vol. 3, pp. 81-90.

Stress is on relation of regional definition and delimitation to purpose. Although the report deals with developmental situations in Latin America, much of the analysis and observation is appropriate to study of regional policies in the U.S.

U.S. Bureau of the Census (B. Berry, author), *METROPOLITAN AREA DEFINITION: A RE-EVALUATION OF CONCEPT AND STATISTICAL PRACTICE* (rev.), Bureau of the Census, Working Paper No. 28., Washington, D.C., 1969.

Short, concise summary of the metropolitan area concept from early history (1910) to a critique of the Standard Metropolitan Statistical Area used in the 1960 Census. The report develops alternative criteria to define a more useful statistical reporting area than the SMSA. Berry's work in Functional Economic Areas and core-periphery traverse analysis is included.

Vining, R., "Delimitation of Economic Areas: Statistical Conceptions in the Study of the Spatial Structure of an Economic System," *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION*, Vol. 48, pp. 44-64, 1953.

Early critique of the Bureau of the Census statistical grouping. Argues against idea that there are "natural" units for economic studies.

Whebell, C.F.J., "Corridors: A Theory of Urban Systems," *ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS*, Vol. 59, No. 1, March 1969, pp. 1-26.

Corridor is defined as a linear system of urban places linked by surface transport. The corridor is shown to be a major type of urban system, currently and historically. Builds theory in first part. Empirical work, taken from a study of Southern Ontario, Canada, is used to illustrate theory.

### Sources for Figures

1. Adapted from O.D. Duncan, et al., METROPOLIS AND REGION, The Johns Hopkins Press, Baltimore, 1960, p. 94.
2. Adapted. Ibid., p. 97.
3. Adapted from B.J.L. Berry, P.G. Goheen, and H. Goldstein, METROPOLITAN AREA DEFINITION: A RE-EVALUATION OF CONCEPT AND STATISTICAL PRACTICE, (revised), U.S. Department of Commerce, Bureau of the Census, Working Paper 28, p. 22, Figure 11.
4. Adapted from NATIONAL TRANSPORTATION ACT: HEARINGS BEFORE THE COMMITTEE ON COMMERCE, 91st Cong., 2d Sess. on S.924 and S.2425, Serial 91-69, U.S. Government Printing Office, Washington, D.C., 1970, p. 196.
5. Adapted. B.J.L. Berry, et al., op. cit., pp. 28-29, Figure 18.
6. Adapted from John Friedmann and John Miller, "The Urban Field," JOURNAL OF THE AMERICAN INSTITUTE OF PLANNERS, Vol. 31, No. 4, November 1965, p. 314.
7. Adapted from THE NEW YORK TIMES, March 23, 1971, p. 1.
8. Adapted from Niles Hansen, RURAL POVERTY AND THE URBAN CRISIS: A STRATEGY FOR REGIONAL DEVELOPMENT, Indiana University Press, Bloomington, 1970.
9. Adapted from Bureau of the Budget, SIMPLIFYING FEDERAL AID TO STATES AND COMMUNITIES, Executive Office of the President, Bureau of the Budget, Washington, D.C., March 1970, p. 6.
10. H.S. Perloff, et al., REGIONS, RESOURCES AND ECONOMIC GROWTH, The Johns Hopkins Press, Baltimore, 1960, p. 5.

## 28 Urban and Regional Imbalances in Economic Development\*

William Alonso

Most developing countries seem to think that they suffer from gigantism of their principal cities, and this view is shared by many experts. It is an ill defined disease. In some cases the worrisomely big city is quite small by comparison to urban areas in other countries, but looms large and is growing rapidly by comparison to the other cities in the country. In a descriptive sense, this phenomenon has been called "primacy" and is often thought to be associated with underdevelopment or the early stages of development. The most common economic argument for calling this concentration excessive is the belief that per capita costs, particularly for infrastructure investment, rise after a certain urban size. However, there is no agreement as to the size at which this occurs; nor, for that matter, is there solid evidence that costs do in fact increase with urban size for a given level of services and facilities.

A companion argument points to the extreme differences in income between the principal cities and the more backward areas, and policies for the development of the less advanced regions are frequently justified on the grounds of economic justice, although it is sometimes also argued that such policies serve to accelerate total national economic development.<sup>1</sup>

Thus, urbanization and regional policies often converge and complement each other. The former tries to steer development away from over-urbanized regions, and the latter tries to attract and promote development in backward regions.<sup>2</sup> This paper will try to show that these complementary policies may often be detrimental to economic growth.

It must be noted, however, that two policy goals are involved in these policies: efficiency and equity. By efficiency is meant, most simply, national economic growth, often measured in terms of per capita national product or a discounted consumption stream. By equity is meant a more equal distribution of income.<sup>3</sup> Other goals are possible and frequently referred to—for instance, occupation of territory for national defense, as in the case of Israel; or the use of regional planning as a technical device to insure the territorial congruence of plans which have been drawn sectorially; or the pleasing of dissatisfied populations to promote political stability. However, most statements of goals may be paraphrased as: "Of course, the

principal goal is efficiency, but many other goals are possible, such as equity." Such statements introduce the miscellaneous category too soon; the equity goal is held frequently enough to be named and opposed to efficiency.<sup>4</sup> After equity, other goals seem to be more randomly held. The point here is that efficiency and equity are two different commonly held goals, and that they may be in conflict. For instance, it is conceivable that the path of fastest economic growth may imply sharp geographic inequalities, concentrating wealth and power in a few advanced centers and condemning backward areas to lengthy periods of poverty. Conversely, policies of regional equalization may slow down the growth of the total economy.

The ambiguity of concepts of regional definition, particularly with respect to the scale of regions, calls for discussion here of an important theoretical development which has recently achieved prominence in several variations, of which the best known is that associated with Francois Perroux and his term "growth poles."<sup>5</sup> This approach transfers most of the problem of inequality from the national to the regional level. It emphasizes the importance of principal regional centers for growth and permits, for instance, policies that seek to equalize income between regions while recognizing the correlation between urbanization and development. Such a strategy has been called "concentrated decentralization."<sup>6</sup> However, it must be noted that, while regional averages may converge, it is perfectly possible that territorial inequalities in income within regions will be greatly increased, and that even if regional *averages* are brought within a common range, no improvement may be obtained in the distribution of income among the population as a whole. More importantly, perhaps, there is no guidance as to the proper size of regions and centers. Thus, a country too finely divided into many regions would be inviting a spreadout and probably inefficient distribution of investment over its territory; on the other hand, a country too grossly divided into few regions might mask over-urbanization or primacy behind regional averages. Since, even if we exclude such as San Marino and Andorra, nations vary in their population by three orders of magnitude, all of a small country may be smaller than a region in a large country, and a pattern which might be viewed as perfectly acceptable dispersal in the large country might be viewed with dismay by the small one as extreme polarization.

This obvious point bears emphasis. The population of a great many developing countries is no greater than that of a good-sized metropolis. Yet, for what may be no more than esthetic preference, most nations, regardless of size, divide their territory into about the same number of planning regions, not far from a range from ten to two dozen. To illustrate by an example which is not extreme, the total population of Chile is about half of that of the New York metropolis. Its total gross national product is comparable to the gross regional product of metropolitan Buffalo, N.Y. Yet regional disparities are one of the principal concerns of Chile, and justly so. Santiago, it appears to be generally held, is becoming too big.

But, from a point of view of international comparison, it is a very modest capital. The principal argument used to demonstrate the excessive size of Santiago is its large size in proportion to total national population and to national urban population. Yet, if a long-range view is taken, and Chile is viewed as part of a Latin American common market, the size of Santiago seems rather small in an array of cities that includes Río, São Paulo, Buenos Aires, and Lima. The point of all of this is that, in the absence of empirically determined scales of relevant sizes of regions and growth poles,<sup>7</sup> these concepts have limited relevance as a guide to policy and action. They have served, however, to prevent the spreading of investment like butter on a piece of bread. The present view is that the national territory is to be covered by distributed centers, like a ham with cloves.

Within the efficiency goal, the complementary policies of urbanization and regional development contain implicit assumptions regarding questions of fact which have not yet been demonstrated. The two principal ones are: (1) urbanization costs per capita (for housing, services, commuting, etc.) rise after a certain population size, and consequently large cities are uneconomic locations for development; and (2) the less developed regions, being relatively virgin territory, must hold opportunities for very high returns. In a sense, these views amount to a simple concept of diminishing returns to scale with size, implying that returns are higher in the less developed regions. If this is true, there appears to be no conflict between the goals of efficiency and equity, since total product will be maximized when all regions are equally developed, or developed in proportion to their natural endowment. Another frequent argument for this point of view is that the development of backward regions serves national development by creating the markets necessary to achieve the necessary demand for efficient scale of production. I shall not address myself to this argument in this paper, except to note that it carries strong implicit assumptions as to the relative income elasticities of demand for the contemplated production mix.

The basic questions remain: are bigger cities less efficient than smaller ones? Are there rich unexploited opportunities in the more backward regions? If the answers to these questions are in the affirmative, policies of counterurbanization and regional equalization are indicated, from the point of view of both efficiency and equity. If the answers are in the negative, efficiency goals will be served by encouraging geographic concentration and regional inequality, and there exists a conflict between the two goals.<sup>8</sup> The answers to these dilemmas are of vital importance. As of today, it may be generalized that virtually every nation that deals with this issue in its national plan has based its policy on the belief that its principal city is too big, regardless of its size. Among the many developed countries which are following policies of relative or absolute dispersal are France, Great Britain, Poland, and the U.S.S.R. Among the underdeveloped countries, Puerto Rico, India, Chile, Venezuela, Ghana, and Egypt are

just a few of many. However, these plans very seldom state with useful clarity their assumptions of fact or their policy alternatives.

First, let us examine the matter of relative efficiency of urban size. The formulation of this issue in effect equates the city to a giant firm. Per capita costs of urbanization are said to decline, to a point, and to rise thereafter. A review of the literature on this issue affords no consensus on the population size at which this turning point is reached. But even if we knew the point of least costs per worker or inhabitant, the issue of optimal size for efficiency would not be answered. To paraphrase Keynes, the engine of business is not economy but profit. That is to say, even if costs rise after a certain point, where productivity is rising faster (by reason of external economies or economies of scale) big cities will yield a greater net return per worker or inhabitant than smaller ones. Thus, even under a condition of rising marginal costs, net marginal product per capita may be not only positive but increasing.

All of these statements must unfortunately be couched in the conditional for, although this is a central issue for the planning of development, there exists an abundance of opinion but a paucity of facts. Considerable evidence conforms to a belief in increasing returns to urban scale. For countries in which data exist, it may be generalized that the larger the city, the higher the per capita income. It must be noted, however, that it has been suggested without proof that a possible explanation for this positive correlation lies not in efficiency but in exploitative situations, based on absentee ownership, political and economic power, perverse flows of funds and skills, and other parasitic factors.

More detailed data exists, to my knowledge, for only two countries, both among the wealthiest. In the cities of the German Federal Republic, per capita regional product rises with population considerably faster than governmental expenditures including welfare.<sup>9</sup> In the United States, per capita income rises strongly and unequivocally with urban size. Governmental expenditures per capita tend to rise, but by far smaller absolute magnitudes, suggesting that, if an "optimal size" does indeed exist, it is far likelier to depend on the productivity-per-capita function than on the cost-per-capita function.<sup>10</sup> It may be noted that none of half a dozen price indices comparing living costs among urban areas shows relation to urban size.

In brief, there is no basis for the belief that primacy or overurbanization *per se* is detrimental to the efficiency goal of economic development. There are good grounds for believing in increasing returns to urban size.<sup>11</sup>

Neither has there been empirical demonstration that a policy of regional equalization is consonant with rapid economic development. If the backward regions are believed to contain unexploited opportunities, and there is a general expectation of decreasing returns to capital in any one region, then classical economics would indicate that a distribution of investment proportional to resources and population would result in the

fastest national growth. However, classical location theory is ill suited to the conditions of developing nations. It is based on perfect knowledge, predictability, mobility of factors, inexhaustible entrepreneurship, and, to a large extent, a fully developed transportation network. None of these conditions obtain in developing countries. A full adaptation of location theory to the conditions of developing countries is beyond this brief paper, but a few illustrations may be given.

The transport network is typically not a full lattice covering the national territory, but rather a tree-shape in which the branches converge on the great port cities. These cities therefore, although they appear peripheral on an ordinary map, are the most central (accessible) in functional terms to the nation as a whole. Further, the frequent intervention of import and export of products and materials in the productive processes of developing nations gives extraordinary importance to these break-of-bulk points. Neither is information evenly available. The more remote the region, the less is known about it. This includes local customs, the chemical and physical properties of local materials, the cost and timing of operations, the quality and seasonal patterns of water, and a thousand other things, all of which add to the uncertainty of the investment decision, whether private or public, and thus require higher margins of return to justify investment. These uncertainties are, of course, in addition to the certainties of poorer communications, slower deliveries, incompetence or absence of supporting services, etc. The uncertainties add to the risks, and the certainties to the costs, and both lead to a need for large and clear advantages to justify locations outside of the developed centers.

A further important factor in the location of industry in developing countries which is neglected in classical theory is that of entrepreneurship and technical skills. Universally, managers and technicians prefer to live in the major cities for obvious personal reasons, and inducements to get them to go to less developed areas are not only likely to be expensive, but, even when successful, may produce only second-raters. Needless to say, especially in developing countries, the quality of management and the competence of technicians is likely to be the most important single factor for an enterprise. This is because the uniqueness and fluidity of each situation call for the highest order of personal ability and energy, since precedents and the supporting institutional matrix are deficient. Further, especially in the area of management, conditions in developing countries change rapidly and often unpredictably, as ministries are reorganized, regulations are changed, the cost and availability of funds fluctuate, relative factor prices vary suddenly, etc. In contrast with developed countries, where comparable changes are usually less frequent and less radical, and where information is easily available in a depersonalized form through printed documents or efficient bureaucracies with clear spheres of authority, in developing countries personal contacts, rumor, influence, and unfortunately, in many cases, extralegal procedures are of far greater

importance. In short, information is likely to be vitally needed on short notice because of the rapidity and frequency of changing circumstances, and, since the sources and modes of this information are unperiodic, unpredictable, and require cultivation of sources, evaluation of intangibles, and delicate negotiations which can only be conducted through personal, face-to-face contacts, they exercise a tremendous pressure toward spatial concentration. Thus, a hard-pencil advantage of a few percentages in one or another cost, which might be decisive in a developed country, is likely to be washed out in the interest of the flexibility and adaptability of a central location in a developing country.

These pressures toward location in the large cities are only slightly less likely to operate in the case of state enterprises than in the case of private ones as long as the efficiency goal is the principal one considered, and the equity goal is seldom recognized explicitly. Even within the efficiency goal, the procedures for accounting for most governmental projects generally are, as are those for private enterprises, based on money costs to the enterprise rather than on net costs to the economy. Thus, it is usually ignored that the price of materials which include a tax or a duty is different from the point of view of the economy than from the point of view of the project. Similarly, most land rent would not represent a cost from the point of view of the national economy, while it would undoubtedly be a cost on the books of the project. In fact, certain types of rent would constitute an addition to national wealth. The project will see most of its cost as average costs, but they will affect the national economy as marginal costs. Such differences between money costs and economic costs and benefits could be handled to some degree by such emerging techniques as cost-benefit analysis, but they very rarely are.

Yet other considerations *within* the efficiency goal may result in different location strategies between nationally planned growth and incremental project-by-project growth, whether private or public. The first of these is the issue of external economies. In some specific cases vertical or horizontal linkages can be handled by relatively well established methods such as input-output or industrial complex analysis, but in the majority of the cases the important relations elude us, and we must refer again to the macroeconomic consideration of efficient urban size. A particular project or group of projects, each of which may be of little interest in itself, may serve to bring a particular region or secondary urban area to a critical mass sufficient to start off vigorous and self-sustaining growth and may thus be justified in terms of induced external economies. We have only vague and anecdotal ideas as to the conditions for such growth: there must be a sufficient variety of activities that unexpected relations and complementarities can be discovered and efficient specialization is possible; the pool of skills must be sufficiently large and varied to permit change and adaptability; life for the upper classes must be sufficiently varied and interesting to keep the best men from fleeing provincial



dullness; local markets must become large enough to attract their own suppliers; the demand for some services must become large enough to justify some major forms of infrastructure such as large airports or specialized financial services. All of these things are relatively well known, but we do not know much about the critical sizes or thresholds involved, especially for the subtler forms of human interaction, so that we cannot measure or predict the contributions to externalities of particular projects, even in cases in which these would be decisive in the calculation of costs and benefits.

Other considerations may indicate a divergence between the national interest as a whole and incremental growth which considers each project individually. The asymmetry of risks and costs and the motivations of managers or project directors may quite sensibly lead to a national policy that discourages risks and encourages central locations.<sup>12</sup> But a nation that contemplates a multiplicity of projects as well as external benefits might be willing to accept occasional failures, and to encourage by subsidies, insurance, or direct intervention groups of projects some of which would be rejected if considered by themselves. The principle is similar to that of health insurance, which, by distributing skewed risks, permits lower per capita investment for the same degree of coverage.<sup>13</sup>

A final set of considerations, which may be termed structural changes, may be considered separately, although they are little understood and are hard to distinguish from the previous considerations. By structural changes are meant the long-range changes in the macroeconomic landscape of a nation. Let me illustrate by a grand example. The pre-Colombian cities of Latin America were, in large degree, central place cities oriented to the national territories.<sup>14</sup> During the colonial centuries into the present, the principal cities of Latin America have been port cities, oriented to the break-of-bulk and transshipment functions that characterize colonial economies. In the last decade, as Latin America has tried to struggle out of the colonialism of economic dependence, we have witnessed the emergence, after half a millennium, of a new round of internal cities: Santo Tome in Venezuela, Brasilia in Brazil, several newly invigorated cities such as Cordoba in Argentina, the trans-Andean experiments of Peru. Other structural changes which may be expected include the reorientation of transportation systems from the port-centered fan-shapes of colonial economies to the fuller reticulation of national development, and the rise of some centers and the decline of others as land reform and other changes in agriculture take hold and as the modes, speed, and costs of transportation change the geographic range of central places and complementary activities. Massive movements of population are a certainty, and the rapid evolution of social and economic behavior will bring about changes in quantity so significant that they will amount to changes in kind. The most familiar example of such transformations is that of nations which, from having a vast majority of their population in rural occupations, have

passed to having a vast majority living in cities. These nations bear witness to its being a change in kind.

My point is that a long-run view of development in a nation must anticipate that fundamental transformations will occur, although their precise form may not be known. Project-by-project analysis must of necessity assume the stability of certain parameters and be essentially incremental. This is amply demonstrated by the most advanced econometric techniques currently in use: they use rates of change or proportionalities which are in effect partial differentials of a first or second degree of models which are essentially static or which grow by expansion rather than by structural transformation or redefinition of categories. This may seem abstract and as imperfectly understood by this writer as by the readers, and therefore further illustration may be helpful. The manufacture of many products will pass from an artisan to an industrial mode of production; many nations, especially in Africa, will pass from a cultural-ethnic solidarity among segments of their population (i.e., tribal) to an organization along territorial (i.e., provinces or regions), sectoral (i.e., industries), or class lines. Acceptance of style by consumers will pass from a basis on custom and tradition to a basis on novelty or other types of symbolism. I believe that it can be said with fairness that our best quantitative techniques take minimal notice of these fundamental transformations. Perhaps this issue is one of the reasons for the division in so many countries between economic and physical regional planners. The former are more sophisticated and more accurate but deal with short perspectives of time, perhaps five years, while the latter, often more romantic and intuitive, deal with longer perspectives, on the order of twenty-five to forty years.

In brief, location theory adapted to developing countries would indicate that the efficiency goal is best served by a policy that permits concentration, at least in the short run. Long-range considerations of the changing macroeconomic landscape may modify this position to one of concentrated decentralization, but this is not clear. Below we shall discuss some recent theory and empirical evidence of such changes in the economic landscape of developing countries. These suggest that concentration is typical of the take-off stage of development, and that equalization takes place as the economy matures.

Albert Hirschman<sup>15</sup> and Gunnar Myrdal<sup>16</sup> arrived at fairly similar models of spatial polarization of economies in the process of development. In the early stages of development, the advantage lies with the developed centers, which enjoy the existence of overhead facilities, external economies, political power, spatial preferences of the decision-makers, immigration of the more vigorous and educated elements from the underdeveloped regions, flows of funds from the land-wealthy in the hinterlands to the financial markets in the cities, and a variety of other factors. These factors lead to polarization, that is to say, to concentration in the large cities and increases in the differences of regional incomes.<sup>17</sup> After a certain point,

however, certain "trickling-down effects" come into prominence. The spread of literacy and bureaucratic practices improve knowledge in and about the backward areas; the opening of transportation routes to reach these areas as markets for the developed centers also opens them as possible locations for productive activities; universal education and standardization of all aspects of life permit an integration of the space-economy and, by making externalities more nearly comparable everywhere, make the more distant opportunities more accessible and more interesting for development. Thus, in this view, in the early stages of development there will be increasing disparity between developed and underdeveloped regions, but there will be a tendency toward equalization as the economy reaches maturity.<sup>18</sup>

Two recent empirical studies support this view. Williamson's<sup>19</sup> findings may be summarized as follows: (1) regional disparities are greater in less developed countries and smaller in the more developed; (2) over time, regional disparities increase in the less developed countries and decrease in the more developed. These findings too suggest that regional inequality, if plotted against economic development, would result in a bell-shaped curve, with some peak being reached at the transition from the take-off to the mature stage. Similarly, El Shaks<sup>20</sup> finds that a cross-section of the world's nations results in a near-normal curve of primacy on economic development. That is to say, primacy is rare in very underdeveloped countries, rises during the take-off stage, and decreases thereafter. His time-series studies of developed countries also support this view, although not as neatly as the cross-section data. To avoid confusion, it must be stressed that this does not refer to suburbanization of large cities, which is an internal reorganization of the urban mass at a totally different geographic scale. Neither does it mean that the large urban areas break up and become smaller. Far from it, the largest cities continue to grow in the mature stage of the economy, but secondary centers grow faster.

It would appear, then, from Hirschman's theoretical approach and Williamson's and El Shaks' empirical observations, that polarization, regional inequality, or primacy are normal aspects of the early stages of development, corrected by natural processes (a form of negative feedback) with the achievement of development. Primacy, overurbanization, and gigantism are not diseases but growing pains. It is tempting to conclude that national policies need not concern themselves with fighting further urbanization in the principal cities or with regional balance as such, and that developing nations should concentrate on speeding the growth of national product. An invisible hand may be at work, and, given world enough and time, it may reconcile the efficiency and the equity goals. But there exist the realities of political pressures and the equally real altruistic desire among many for early equalization, and some difficult questions remain unanswered by this analysis.

Given that the wholehearted pursuit of efficiency goals will in due

time result in satisfying the equity goals, can we be sure that the opposite does not hold? If regional equity is in some measure actively pursued for its own sake, may it not serve to accelerate national growth? This may be the case if a series of decisions made on short-term bases toward efficiency miss out on the opportunities of the changing macrogeography of development. We argued, for instance, that decisions might be quite different if probabilistic studies of costs and benefit are made of groups of projects rather than for single projects. We also argued that decisions based on expected secular changes of what may be termed the macrospatial aspects of the economy might differ substantially from decisions made within the context of five-year plans. Or to put it another way, viewing existing and possible centers of activity in terms of the classical theory of the firm, chains of decisions on alternative investments based in each case on the contemporary marginal cost and product curves may lead to decisions quite different from those based on long-run curves. The answers to these questions remain a matter of guesswork, emotion, and opinion for lack of factual information, although in my opinion the burden of proof should rest on the theories which favor dispersal, since existing evidence and some existing theory point to concentration during the take-off stages as at least one possible path to development.

But there is another class of policy problems which remain unanswered for lack of conceptual tools. Whereas the efficiency goal has a relatively objective measure through the concepts of national income, and particular choices over time may to some extent be compared through discount rates, no such conceptual tools are available for the equity goal. As we have shown, even the grossest measures of equity depend on the grain or scale of regional definition. More basically, given two alternative distributions of income, we often have no way of determining which is more equitable, particularly if more than two regions are involved.<sup>21</sup> There are no accepted ways of comparing successive distributions of regional income over time comparable to the techniques for discounting future income. Further, since the goals of equity and efficiency are usually held simultaneously, the problem of how much growth in national product would be surrendered for a given amount of equity (if it were measurable) is one that plagues the planning and political functions of nearly every country. It is no wonder that the advocates of both centralization and decentralization hope that their approach will satisfy both goals simultaneously, and that this difficult choice may be avoided. It may be hazarded that, should a simple and intuitively satisfactory numerical measure of equity be devised, it would do as much to revolutionize planning thought and techniques as has the introduction of the concept of rates of growth in national income since the Second World War. This would be particularly true if this measure of equity were cognate with the measure of efficiency, and the two could be compared.

To summarize: (1) it appears that a policy that pursues concentration

is a possible path to satisfy the efficiency goal; (2) such a policy may be in agreement with the equity goal, especially in the long run. However, (3) we do not know whether long-run structural changes make possible some form of decentralization policy that satisfies the efficiency goal better, while recognizing earlier the equity goal. Finally, (4) there are no good quantitative criteria for comparing alternative regional distributions with respect to equity, especially if a time dimension is added. Thus, there is great need for factual knowledge to determine what alternatives are open and for the development of conceptual tools for choosing among these alternatives.

\* This work was conducted as part of the Harvard Regional Economic Development Program, supported by a grant from the Economic Development Administration of the Department of Commerce. It also reflects earlier work by the author at the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University.

1 Sometimes one encounters advocacy of what may be called a naïve form of balanced regional growth, which holds in effect that all parts of the nation must grow at the same rate. It has recently been pointed out that the "big-push," sectorially balanced growth strategy of Rosenstein-Rodan and others, which stresses the complementarity of certain sectors, is likely to result in sharp differences in regional rates of growth, with a bias toward the more developed centers, thus leading to greater regional inequality. See W. F. Ilchman and R. C. Bhargava, "Balanced Thought and Economic Growth," *Economic Development and Cultural Change*, Vol. 14, No. 4 (July 1966), pp. 385-99.

2 In developed countries, depressed regions are more often the object of regional policy than backward regions. The frontier region, which is of importance principally in developing countries, will not be discussed in this brief paper. Neither will the relative priorities of agricultural and other types of development be discussed.

3 This issue, of course, is often taken up in terms of inequities of income among classes or sectors, but our concern here will be with interregional comparisons of income.

4 These two goals are discussed abundantly if abstractly in regard to the meaning of the Pareto optimum in welfare economies.

5 See, for instance, F. Perroux, "Note sur la notion de 'pôle de croissance,'" *Economie appliquée* (January-June 1953).

6 L. Rodwin, "Metropolitan Policy for Developing Areas," *Daedalus* (Winter 1961).

7 There is at this time very little sense as to the relevant dimensions of such scales. It is likely that they may have to combine population, some measure of size of economy, and distance or area. Thus, in the case of Chile, its long, thin shape probably augments regional differentiation and its perception.

8 This issue is slightly more complex and admits of other alternatives which will be treated in a subsequent paper. For instance, marginal productivity of labor in backward areas may be below the average product in those areas. In that case, migration from the country to the city would in fact raise income per capita in the backward areas. This may occur, for instance, in cases of overpopulation and too small farms, or of unemployment (sometimes disguised) in rural areas. If the marginal productivity of such people is greater in the large cities than in the hinterland, there exists no conflict between concentration and income equalization. It will be obvious to the reader that I am also ignoring the possibility of transfer payments from the rich to the poor regions as a means of avoiding the dilemma.

9 Based on *Statistisches Jahrbuch Deutschen Gemeinden* (52) (Braunschweig: Waisenhaus Buchdruckerei). The data is for 1964. I thank H. Schlegel for directing me to this source.

10 For instance, local government expenditures per capita in U.S. counties in 1957 rose from about \$120 for populations of about 50,000 to about \$200 for populations of over 1,000,000. H. J. Schmandt and G. R. Stephens, "Local Government Expenditure Patterns in the United States," *Land Economics* (November 1963). But median family income rose by \$1,100 over the same range. L. F. Schnore, "Some Correlates of Urban Size: A Replication," *American Journal of Sociology* (September 1963). Even if we disregard the greater range and quality of services in large cities, the rise in family income translated into per capita terms is four times as large as the rise in government expenditures. It may be noted that in the United States a great deal of the literature concerns itself with the problems of governmental costs by municipal subdivisions of metropolitan areas, and that in that literature a distinction is drawn between the city (a municipality) and a metropolitan area. In this paper, it is obvious, the city is used to mean the whole of an urban area.

11 One argument sometimes advanced, but to my knowledge not fully developed, is that the rate rather than the level of urbanization is uneconomic. In other words, it is not the size of cities that is costly, but the rapidity with which they are growing. Needless to say, no data exists showing whether this is true. Theoretically, one may argue that many lumpy elements of infrastructure could be more efficiently provided under conditions of rapid growth, with shorter periods of idle capacity. Some interesting work on the engineering aspects of this, termed a "threshold theory," are reported for Polish cities by B. Malisz, in J. Fisher, ed., *City and Regional Planning in Poland* (Ithaca: Cornell University Press, 1966). On the other hand, the rapidity of the growth

may strain the capacity of certain sectors (principally construction, education, civil order) and result in costly or bad solutions.

12 On this point, see W. Isard and T. A. Reiner, "Aspects of Decision-Making Theory and Regional Science," *Papers and Proceedings of the Regional Science Association*, Vol. 9 (1962).

13 This point bears a close relation to the discussion of a "social rate of discount" among welfare economists. However, their arguments are based on the different time preferences of society and individual investment projects. The argument here is based on the effects of large numbers on probabilities and decision strategies and is similar to the "principle of massed reserves" in the standard discussions of economies of scale and external economies. In the United States this principle has been successfully applied in housing through the federal mortgage insurance program, and it is now being tentatively applied to the field of foreign investment.

14 See J. E. Hardoy, *Las Ciudades Precolombinas* (Buenos Aires: Ediciones Infinito, 1964).

15 A. O. Hirschman, *The Strategy of Economic Development* (New Haven: Yale University Press, 1958), Ch. 10.

16 G. Myrdal, *Rich Lands and Poor* (New York: Harper, 1957).

17 These arguments and the related ones presented in this paper are new to a large extent only in emphasis and in their policy orientation. For instance, Edward Ullman, in "Regional Development and Concentration," *Papers and Proceedings of the Regional Science Association*, Vol. 4 (1958), states succinctly that large cities grow because they enjoy externalities and economies of scale and because they are centers of decision-making and innovation. Distant areas lack these advantages and, being remote, tend to suffer from higher transportation costs.

18 Myrdal is more pessimistic than Hirschman about this eventual convergence and stops his analysis with what he terms the vicious circle of backwash effects, which are equivalent to Hirschman's polarization.

19 J. G. Williamson, "Regional Inequality and the Process of National Development: A Description of the Patterns," *Economic Development and Cultural Change*, Vol. 13, No. 4, Part II (July 1965).

20 S. El Shaks, "Development, Primacy, and the Structure of Cities," unpublished Ph.D. dissertation, Harvard University, 1965. It should be noted that El Shaks achieves a significant theoretical improvement in the definition of primacy by the construction of an index which considers the distribution of urban sizes as a whole. This may explain the difference between El Shaks' findings and those of B. Berry, who finds no pattern. See N. Ginsburg and

B. L. Berry, *Atlas of Economic Development* (Chicago: University of Chicago Press, 1961), p. 36.

21 The case of Italy's Mezzogiorno is instructive here. In its early stages the policy was a simple one of equalization between North and South. It was soon abandoned, and instead a policy of achieving sufficient scale for self-sustaining development in the South was substituted. In other words, a subnational efficiency goal was substituted for a national equity goal. See, for instance, H. B. Chenery, "Development Policies for Southern Italy," *Quarterly Journal of Economics*, Vol. 76 (November 1962).



# 29 Problems, Purposes, and Implicit Policies for a National Strategy of Urbanization

William Alonso

## INTRODUCTION

The course of our national domestic history in the first 12 decades of our nation was largely the settling of the continent, and national policy was oriented to this purpose. But our efforts during this century have not had such sweep. Rather, they have focused on solving the problems of distressed areas. At best, the problem-oriented approach has yielded mixed results in such cases as TVA, Appalachia, and the central ghettos of our large urban areas. Only in recent years has there been any attempt to develop a national policy for the overall distribution of population and economic activity. Unfortunately, although the need for a policy approach is strongly sensed, the purposes and the methods of achieving an overall policy are still poorly defined.

Many of the current policy proposals are naive either because they do not recognize the realities of the working of the system or because they propose geometrics of population distribution without relating them to national purposes. Why should a policy of developing middle-sized alternate growth centers be better than a policy of steering growth towards very small centers or towards very big centers? Would one policy or another improve material well-being, diminish inequality, conserve the environment, or provide more satisfactory ways of life? Clearly, choosing among such policies, matters of feasibility aside, would depend on understanding their implications in terms of our national purposes. But, if this is too ambitious, policies should at least specify concretely the problems to which they are addressed.

## PROBLEMS IN POPULATION DISTRIBUTION POLICIES

Any policy of territorial distribution that amounts to more than window dressing will involve vast sums of money and affect all aspects of our national life at a scale which is not sufficiently realized. The programs involved would dwarf such current efforts as the space program, or such proposed ones as a national welfare system or a national health insurance. Further, popula-

tion distribution is the result of a slow and steady process, so that any effective program will have to be sustained for decades, if not forever. Thus, it is very difficult to determine how confident we are that one or another policy will meet a general purpose or solve a particular problem before placing a heavy bet on it. It may be ineffective or, worse, counterproductive. Geographic radicalism seems misplaced when we are uncertain of the purposes, of the processes, and of the effectiveness of the instruments.

Our low level of understanding may be illustrated by some facts of geographic demography which run counter to popular stereotypes—our metropolitan areas are awash with urbanizing migrants, and all population is gravitating towards the largest metropolises. Rather than this, facts are that metropolitan areas grow almost exclusively by natural increase. In the last decade, urbanization in rural and small town populations was a very small fraction of metropolitan growth (about 6 percent), and international migration was three times as large a factor. In metropolitan areas (over 2,000,000 population) the lag in terms of net migration, with two migrants per thousand population per year. This is half of the metropolitan average. Far from being in decline, metropolitan areas between 200,000 and 250,000 population are among the fastest growing areas, as are those from 750,000 to 1,000,000. Indeed, one out of three metropolitan areas in the 50,000 to 200,000 class is attracting migrants at twice the overall metropolitan rate, compared to one in eight at the turn of the century. Least realized, perhaps, is that nearly half of the metropolitan areas were net losers through migration; this proportion held for those above as for those below 250,000, and indeed for the 20 largest.

Of course, natural increase has insured that 50 percent of metropolitan areas have continued to grow. But it is quite likely that a continued decrease in the birth rate (brought about by changing life styles, public policy, and the climate of opinion), together with an expected further decrease in rural to urban migration, may have totally unanticipated results by 1980: Some one-third of our metropolitan areas (not just the central cities) will be losing population without benefit.

Reprinted from the U.S. Commission on Population Growth and the American Future, *Research Reports*, vol. 5, *Population Distribution and Policy*, ed. Sara Mills Mazie (Washington: Government Printing Office, 1972), pp. 635-647.

of any national policy of deconcentration. Today, as the issue of population is discovered, attention focuses almost exclusively on the problems of urban growth. But, if nothing else, we have been living with such problems for three centuries. We know virtually nothing about the problems of population decline in large urban areas, except that they are different than the problems of growth. One may speculate about the fiscal situation of local governments, about a shrinkage and increased instability of growth-oriented sectors such as construction, about the consequences of changes in the housing market for various income groups. Not all of the consequences would be bad, of course, but many would. Yet the neglect of such a visible and probable issue illustrates the primitive level of most discussions of territorial policy. It clearly shows that we do not yet have a full understanding of population distribution as a system; consequently, it is unlikely that we can devise policies which are effective yet do not have major, and probably unwanted, unanticipated consequences.

The previous example meant to show our poor understanding of the spatial workings of the social system; the next is an illustration of how poorly we have thought through our definition of problems. Among those who suggest that our big urban areas have grown too big, it is often argued simultaneously that (1) the big cities are so big and so costly that they are inefficient producers, and (2) the continued gravitation of population to these large centers increases the disparity in material welfare between them and the less developed areas. Although under very special circumstances both statements might be true, a moment's thought will show the near contradiction of the two propositions. If the big urban areas have grown beyond efficiency, their citizen's material welfare would be declining; thus, we would be headed towards reduced interregional inequalities, albeit a rather soreheaded reduction. If both propositions cannot be true, which then is the circumstance and which the purpose of a deconcentration policy?

In brief, the point is that population distribution is the territorial aspect of a highly connected and interdependent social system, and that local variations in welfare and productivity are also aspects of this larger system reflected upon geographic space. To intervene in such a system we need first-rate policy analysis based on an understanding of the system, of the options, and of our purposes. But we lack such understanding and, consequently, the capacity to generate credible master plans. Rather, population distribution policies must be generated over time as the outcome of a social learning process. The felt need for such policies at present arises from a large but finite number of pressing problems

which are believed interrelated. As these problems become better understood (and often redefined), as their interrelations become more traceable, and as our social goals are clarified, we then hopefully will progress from problem-oriented programs to system-oriented policies—from cures for ills to the management of health—based on greater insight into the positive purposes that shape such policy rather than on the negative purposes of most of today's programs. At this time, we have begun to recognize that the problems approach is not enough, but we are not yet able to go much further.

The list of problems which has triggered our search for national policy is long but not infinite, and one may begin the search for the agenda or table of contents of that policy through a listing of those problems. They are the problems of size, of growth, and of decline. At the upper end of the scale of size, there are the problems of congestion, pollution, access to open land, fragmentation of jurisdictions, and social and psychological problems such as personal alienation and the lack of responsiveness of institutions to individuals. At the lower end of the scale of size there are the problems of lack of resources, lack of adaptability to change, a narrowed range of social and economic choices, and increasing dependency and loss of self-determination. The problems of growth, as distinct from those of excessive size, include governmental cash-flow crises in paying for additions to the urban plant out of proportion to the existing population base; the disruption of traffic and land-use arising from the successive installation of major new urban elements; the strains of mutual adjustment of old and new social groups to each other, and of all to a larger urban scale; and the loss of such valued features as agricultural landscapes. There are also the problems of population decline, found in the central cities of metropolitan areas and in many smaller communities. They include the need for consolidation, the depreciation of existing capital stock, the loss of morale, and the welfare problems of a population which is increasingly composed of the old, the uneducated, and the very young. Some problems may have been overlooked in this list, and some of those included might be better labeled, but I believe that there would be no great difficulty in arriving at consensus on some comparable list.

Although all of these problems are associated with population distribution, some of these problems may be best attacked by programs other than territorial ones. For instance, it may be argued that urban poverty and racial problems are problems *in* cities, not problems *of* cities. These problems were created over three centuries ago in rural areas and in small towns. People migrating to

the cities carried these problems with them; therefore, the problems must be solved in the cities since the majority of those suffering poverty and discrimination are there now. There is evidence that, however slowly and painfully, progress is being made in the urban areas and that the movement to the cities is part of the solution. By way of illustration, the proportion of black families below the poverty level decreases steadily with increasing urban size.<sup>1</sup>

Other problems may have their origin in population concentration, yet territorial policies may not be appropriate to solve them. Many forms of pollution, for instance, become matters of concern only in large concentrations of population. Yet if these problems are critical, territorial policies may be largely irrelevant, because (to judge by the experience of other countries which have had vigorous decentralization policies) the rate of population dispersal is so slow that it would be decades before its impact would be felt. The development and adoption of technological solutions and such institutional innovations as pollution taxes would be much quicker and more certain. Thus, if this analysis is correct, such problems as poverty and pollution are best attacked directly, without distractions of a territorial nature.<sup>2</sup>

Other situations—for example, many cases of outmigration—are false territorial problems: Although two-thirds of United States counties had net outmigration in the past decade, substantial outmigration is one of the Federal criteria for defining an area as distressed and for putting into effect programs to try to stem the flow. Yet, the distressed conditions are often the result of a surplus of population in relation to the district's resources and opportunities, and geographic mobility is evidence of the adaptive capacity of the people. In those cases, outmigration may redress the balance of population and resources, and be witness of health, not of pathology. To be sure, problems may result from this, such as a high proportion of population in dependent ages or a need to consolidate schools. However, these are transitional costs which merit attention on the basis of equity and of facilitating the adaptation. In those cases, programs to ease the transition may be justified, but programs to maintain earlier population levels may not.

#### OBJECTIVES FOR A POPULATION DISTRIBUTION POLICY

While in the short run we must address these syndromes of problems, in the longer run territorial policy should be oriented to system objectives. The definition of these objectives will evolve over the years

as we learn more about the system, are confronted with new problems, and become aware of unanticipated constraints and opportunities. But, although a tentative list, four principal objectives for United States policy may be read into current discussions: efficiency, equity, environment, and life style.<sup>3</sup> These objectives are sometimes in conflict, and alternative policies will strike different balances among them. Therefore, social science and policy analysis must clarify the trade-offs among them and political processes must make the choices.

At the national level, the objective of efficiency is most simply that of increasing the aggregate material well-being of the population. A precise definition involves technical complications concerning the discounting of the future as opposed to the present and many ticklish questions of measurement and definition, but it is basically a well understood objective and the one normally measured by conventional cost-benefit analysis. Efficiency enters into territorial policy because considerable evidence shows that alternative patterns of population distribution result in different levels of national product in the short and in the long run.

The objective of equity has to do with the distribution of material consumption—the equality or the fairness with which access to resources and consumption are available to various elements of the population. The issues here deal with who bears the costs and benefits of alternative territorial distributions and of changes in distribution. For instance, it is conceivable that a very high rate of national economic growth is associated with increasing inequality. Or, a policy of equalizing per capita income among regions by industrializing the less developed ones may result in the rapid rise of certain social classes and localities within the latter, so that increasing national interregional equity may result in greater intraregional inequity. In any case, while development of a distressed region is an efficiency objective from the point of view of the region itself, it is an equity objective from the national point of view.

The objectives of efficiency and equity have become well known by now, much discussed in the economics literature if not perfectly understood. But the other two objectives are only now emerging and are much less clearly specified. The environment goal has been the subject of much recent enthusiasm, some of it excessive. It deals on the one hand with certain balances of land, air, and water and of the living things in them. These balances are at least important for the physical comfort and well-being of people and, at the extreme, necessary for their survival. On the other hand, the environmental objective also includes the preservation of certain physical or zoological elements which have value for their

own sake rather than for the sake of people (economists might call them merit goods). The importance of the environmental objective has become clear as air and water (traditional examples of "free goods" to generations of freshmen in economics) have become scarce. Its trade-offs with other goals arise because, by requiring greater amounts of capital and labor and by imposing delays and expensive pre-investment studies, the environmental goal conflicts with the efficiency objective. Similarly, it involves trade-offs with the equity objective by affecting manufacturing and primary industries more than services. Thus, it costs the working class more than the middle class, since the working class depends much more on these sectors for jobs and spends a greater portion of its budget on food and manufactured goods, whose prices are more likely to be raised. Thus, it is no surprise that there are few if any black faces at ecology rallies. In geographic terms, pursuit of this objective will have important consequences. It might help distressed areas by steering economic activity in their direction—giving them a factor-price advantage over more crowded areas which are nearing their carrying capacity. But it may also hurt them if the types of primary and manufacturing activities in which they specialize are sources of widely diffused pollution subject to national regulation. Similarly, their economic progress might conflict with preservation of landscape features.

The human well-being objective is perhaps the vaguest, but its vital importance is obvious to everyone but the narrowest of specialists. The way people are distributed affects their way of life and the way they feel about their lives. But our knowledge of how it does this is at best diffuse, wrapped in rhetoric and myth. Much of the present debate centers on the size of cities. According to one view, big cities impose role-segmented contacts on people and keep them from knowing each other as whole persons. The scale and impersonality of the city keep people from understanding the forces that affect their destinies so that they fall victim to alienation and *anomie*. Smaller places, by contrast, provide a single locus for home, school, job, shop, recreation, and civic activities, and thus afford deep and enduring relationships in a comprehensible environment in which the individual may participate and exert some control. But this view is also opposed by many of our most perceptive writers and sociologists who present a picture that does not square with the equation of the big city and alienation, and that makes smaller places appear stifling. It would appear that some people can lead full and rewarding lives in either kind of place, some in one but not the other, and finally, it must be feared, some people's lives will be unsatisfactory in either.

Given the poor definition of this objective and the obvious fact that no single way of life will be the good life for everyone, or even for one person at all times, it is clear that this objective calls for the provision of choices and for freedom to choose. Yet balancing these choices with the other objectives may be bitter and difficult: Do people who prefer to remain in unproductive places have a claim, under the equity criterion, to draw on the resources of the more productive? If Americans love a way of life that consists of low density suburbs, lower density second homes, and multiple car ownership, how is this to be balanced against the environmental objective?

This list of general objectives is fragile in its ambitions. Its principal purpose is to stress that the objectives of territorial policy are aspects of the common set of national objectives on which all national policy is based, however imperfectly. The very tentativeness of this listing, together with the uncertainty of our policy instruments, must be matched against any proposed grand policy design. If the current population distribution patterns are to be changed by reversing the present trends of population flows, can such policy be clearly related to these or comparable objectives? Further, if we choose policy objectives, do we know what programs will achieve them? The answer to both questions is negative. Although territorial policies should begin to be outlined now, this should be done with greater sobriety than is presently the fashion. Recognition of the length and uncertainty of the undertaking makes clear that what is needed is not a facile master plan but the design of an evolutionary process of social learning, involving fundamental and applied research, citizen participation, and institutional approaches that favor experimentation without excessive commitment and learning from mistakes as well as successes. This point will be picked up again at the conclusion of this paper.

#### EXPLICIT POPULATION DISTRIBUTION POLICIES

It is a curious paradox that most present explicitly territorial policies are thought to be ineffective, while it appears that many other policies and programs, whose intent was not originally territorial, powerfully affect the distribution of population and economic activity. This paradox may be explained by returning to the image of the national socio-economic system as a multidimensional one, where the territorial distribution is merely one perspective—a projection upon geographic space. Programs that try to rearrange the system's

elements directly on the geographic projection will typically underestimate the forces and misjudge directions of effects along other dimensions of the system, and thus be ineffective. Sometimes, indeed, these connections along other dimensions bring about consequences quite different from those intended, as when programs aimed at improving conditions in an economically distressed area through the modernization of industry result in lowered employment by driving out marginal enterprises and substituting capital-intensive technologies.

Similarly, programs that are adopted for reasons that have nothing to do with population distribution may have strong consequences on that dimension. In the pages below, several illustrations of this will be outlined. By analogy, it is as if engineers involved in the design of the production flow, or the elevation or cross-section of a plant, paid no heed to the consequences of their decisions on the ground plan and circulation of personnel, while those engineers concerned with these latter paid no particular attention to the former. While standard engineering practice makes for a fair amount of coordination of these aspects before a plant is built, the level of coordination in national policy is much lower; the actual results are that the cross-section and elevation men are quite unmindful of the ground plan, while those in charge of the ground plan wonder ineffectively why their designs bear so little fruit.

The general ineffectiveness of direct Federal efforts to affect the location of population and economic activity is documented by a recent study,<sup>4</sup> which included a detailed analysis of 42 principal Federal assistance programs having potential impact—under the Departments of Agriculture; Commerce; Health, Education and Welfare; Housing and Urban Development; Interior; Labor; and Transportation; and the Office of Economic Opportunity and the Small Business Administration. These programs were ranked on an impact scale from “none” to “heavy.” The evaluation of their interregional effects ranged from “none” to “slight,” with the single exception of the Federal Aid Highway Construction Program, whose impact was judged “heavy.” Similarly, in assessing the impact of these programs on growth center development, only two of the 42 scored as high as “moderate” (Highway Construction and Rural Electrification Loans).

The general conclusions of the study merit extended paraphrasing. It attributes the limited effectiveness of Federal assistance programs as instruments for affecting geographic distribution to the fact that this distribution and its trends are the result of broad economic forces in the private sector. The Federal programs are neither

designed, administered, nor funded to achieve significant impact, and are further hampered by the unwieldiness of the structure and process of the Federal system. It concludes that (1) aid to business investment (principally loan guarantees, and plant, equipment, and loan subsidies) has the greatest and most immediate effect; (2) aid to public facilities and infrastructures is important for removing barriers for development, but does not stimulate the process initially; and (3) investment in human resources has little direct or demonstrable economic impact. These aid programs have their greatest impact within metropolitan areas, but even here they affect business investment most strongly and population distribution least.

After a less detailed survey of other areas of Federal activity, the study suggests that these are potentially more powerful factors than the assistance programs. It considers the credit system; the regulation of economic activity; the procurement of goods, services, and research and development; and the provision of infrastructure (principally water resource projects). It concludes that their geographic outcomes, although significant, are largely unintentional and therefore unlikely to coincide with deliberate national territorial policy.

These conclusions are supported by an examination of the location of Federal expenditures. Table 1 shows the distribution of these expenditures by functional categories in FY 1968. The concentration ratios are in each case the percentage share of that type of expenditure over the percentage share of population in that type of area. They may be interpreted as the per capita share, where numbers greater than 1.00 indicate more than proportional expenditures.

Overall, expenditures are concentrated in the central cities, the richest and fastest growing counties, and the slowest growing counties. The poorest counties receive the smallest share. Defense contracts and defense payrolls are concentrated in the richest and fastest growing counties. Defense, as the largest budget item, is clearly one of the principal determinants of the location of growth. Although a 1967 Independent Study Board recommended that regional development considerations be taken into account in procurement policies, Congress has insisted that contracts be awarded on the basis of least cost, requesting that the purchases be dispersed only if costs are equally low. Space contracts have been concentrated in rich, fast-growing metropolitan suburbs; these same suburbs, of course, are today suffering the main consequences of cutbacks in this area (the 1968 figures in Table 1 do not show this). Atomic energy, on the other hand, is quite dispersed, concentrated in Economic Development Administration counties, but

not in the poorest. Regional development expenditures, as such, are concentrated in the poorest and slowest-growing (actually population-losing) counties; that is, at least in 1968, aid was going to the neediest areas directly rather than indirectly to growth centers. (This may have changed since then.) Air transportation and mass transit are concentrated in large metropolitan areas, the latter in the central cities. Less predictably, perhaps, ground transportation (principally highways) is low in the richest and fast growing counties, and high in the poorest, rural, and slowest growing counties, while water transportation is concentrated in the central cities of large metropolitan areas. Health expenditures are generally dispersed, as are Retirement Income Security Payments. Other Income Security Payments, however, were concentrated in the poorest counties and central cities of metropolitan areas (the concentration in richest counties is only apparent, arising from an overlap of definitions with central cities).

Federal loans for regional development, like the grants, are concentrated in the poorest and population-losing counties. Business and Economic Opportunity Loans are oriented toward poor, rural, population-losing counties. This is primarily because this category, although including Small Business Administration and Office of Economic Opportunity activities, is dominated (79 percent) by agricultural loans. Housing loans are heavily oriented to the suburbs of metropolitan areas, in contrast with housing grants, which are oriented to central cities.

Although these programs (which are, to a greater or lesser degree, regionally defined) differ to some degree from each other in the types of areas they favor, they sometimes complement each other by tackling the different problems of different kinds of areas. But, on the whole, those programs intended to directly affect geographic distribution make up a relatively small portion of Federal expenditures compared to other programs. These other programs are predicated on purposes which are not territorial, but their results prove not to be random across the national landscape; the explicit non-territorial purpose sometimes carries hidden within it an implicit territorial policy. Further, it will be seen below that direct Federal expenditure is only one type of Federal action that carries implicit territorial policies, and not necessarily the most important.

In the American Federal system, the principal direct forms of local development assistance are viewed as facilitating or inducing development—mainly in depressed areas or central city ghettos. There is at present, in spite of widespread concern with possible excessive development in our urban areas, no set of programs

aimed at discouraging excessive concentration other than some desultory relocation of Federal facilities. Other nations, such as Britain and the Soviet Union, have employed a variety of direct controls on the location of economic activities (and even of population) usually with a purpose of checking the growth of their largest urban concentrations. These areas have continued to grow, however, although they might have grown more without these controls.

The British experience is instructive. In the years after World War II, the objective of their policies was to “decant” the population from the London area. In recent years, the objective has come to be redefined as the guidance of the mode of urbanization of southeast England. The lesson of such experience is that it is extraordinarily difficult to check development where broad economic forces are directing it, even in socialized economies, and that it makes more sense to try to guide these forces than to try to block them.

Several current legislative proposals appear to portend a new emphasis on land-use planning at the state level. This would differ from present direct strategies, which seek to promote development by inducements, in that land-use planning (with the exception of the location of public investment) is primarily a negative control which prevents certain types of development in certain places. Although the form of this type of strategy and its consequences cannot be foretold with any accuracy, it is instructive to keep in mind the experience of the city-planning activity, which has used land-use planning and controls as its principal instrument for many decades. City planning sought to have comprehensive or master planning by covering the totality of the city's territory. But the experience of the past few years has amply demonstrated that this formal geographic comprehensiveness ignored, at great cost, those problems and interrelations which were not easily representable on a map, and that these have proved of overwhelming importance. Thus, although statewide land-use planning may have a role to play, it will only be an aspect of comprehensive regional planning in this fuller sense. This is especially true with respect to urban activities in our urban society. Land-use planning is most useful for those activities which use land extensively. Such activities appear as areas on a land-use plan. But at that scale, urban activities, which take up only about one percent of the national territory, are so small as to be practically dots on the map, and their interrelations will be along dimensions of which adjacency is only one. Regional planning would need to take these other dimensions into account, and they would not appear in the land-use plan.

**Table 1.—Population Shares of Diverse Types of Areas and Federal Expenditures Concentration Ratio by Functional Categories, FY 1968**

|   | % of<br>total<br>outlay <sup>a</sup> | Poorest<br>counties | Richest<br>counties | Slowest<br>growing<br>counties <sup>b</sup> | Fastest<br>growing<br>counties | Metropolitan areas<br>> 1,000,000 | Metropolitan areas<br>< 1,000,000 | Non-SMSA<br>urban<br>counties | Rural<br>counties | Central<br>cities | Suburbs | EDA<br>counties |
|---|--------------------------------------|---------------------|---------------------|---|--------------------------------|-----------------------------------|-----------------------------------|-------------------------------|-------------------|-------------------|---------|-----------------|
| <b>Regional</b>                           |                                      |                     |                     |   |                                |                                   |                                   |                               |                   |                   |         |                 |
| Agri. Program NEC                         | 3.5                                  | 1.43                | 1.65                | 3.07  | 0.42                           | 0.67                              | 0.56                              | 1.38                          | 1.96              | 1.47              | 0.23    | 0.95            |
| Natural Resources                         | 2.5                                  | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Air Transportation                        | 0.5                                  | 0.18                | 1.62                | 0.45  | 0.75                           | 1.49                              | 1.03                              | 0.41                          | 0.41              | 2.39              | 0.49    | 1.18            |
| Water Trans.                              | 0.5                                  | 0.11                | 4.28                | 3.54  | 1.04                           | 1.86                              | 0.52                              | 0.96                          | 0.22              | 4.33              | 1.04    | 1.32            |
| Urban Mass Transit                        | *                                    | 0.00                | 4.37                | 3.37  | 0.13                           | 2.19                              | 0.60                              | 0.03                          | 0.01              | 3.22              | 0.54    | 1.36            |
| Other Ground Trans.                       | 2.4                                  | 1.37                | 0.64                | 1.25  | 0.87                           | 0.73                              | 1.04                              | 0.94                          | 1.43              | 0.77              | 0.62    | 1.07            |
| Regional Dev.                             | 0.2                                  | 4.05                | 0.23                | 2.92  | 0.82                           | 0.38                              | 0.40                              | 1.23                          | 2.74              | 0.89              | 0.05    | 2.07            |
| Housing and Community<br>Aids             | 1.2                                  | 0.70                | 1.42                | 1.44  | 0.59                           | 1.27                              | 1.07                              | 0.77                          | 0.57              | 2.21              | 0.55    | 1.20            |
| Air Pollution Control                     | *                                    | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Water Pollution Control                   | 0.1                                  | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Bus. & Farm. Econ. Oppor-<br>tunity Loans | 2.1                                  | 1.86                | 0.24                | 3.23  | 0.38                           | 0.13                              | 0.39                              | 1.94                          | 2.82              | 0.22              | 0.17    | 0.87            |
| Reg. Develop. Loans                       | 0.4                                  | 3.71                | 0.13                | 2.85  | 0.45                           | 0.11                              | 0.30                              | 1.60                          | 3.14              | 0.18              | 0.19    | 1.32            |
| <b>Mixed regional</b>                     |                                      |                     |                     |   |                                |                                   |                                   |                               |                   |                   |         |                 |
| Recreation                                | 0.1                                  | 0.82                | 0.51                | 3.71  | 0.79                           | 0.73                              | 0.42                              | 0.85                          | 2.31              | 1.71              | 0.57    | 1.81            |
| Business Programs                         | 0.3                                  | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Hous. Mortgage Market                     | *                                    | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Elem. & Sec. Ed.                          | 1.2                                  | 1.52                | 0.85                | 1.14  | 0.96                           | 0.73                              | 1.26                              | 0.96                          | 1.13              | 1.38              | 0.68    | 1.03            |
| Vocational Ed. & Manpower<br>Training     | 0.5                                  | 0.59                | 0.88                | 1.20  | 0.79                           | 0.84                              | 1.62                              | 1.24                          | 0.32              | 2.12              | 0.20    | 0.85            |
| Other Ed. & Manpower<br>Aids              | 0.4                                  | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| General Govt.                             | 1.6                                  | ...                 | ...                 | ...   | ...                            | ...                               | ...                               | ...                           | ...               | ...               | ...     | ...             |
| Housing Loans                             | 6.9                                  | 0.40                | 1.10                | 0.49  | 2.06                           | 1.24                              | 1.23                              | 0.68                          | 0.44              | 1.03              | 1.23    | 1.00            |

|                            |      |      |      |      |      |      |      |      |      |      |      |
|----------------------------|------|------|------|------|------|------|------|------|------|------|------|
| Defense & Science          |      |      |      |      |      |      |      |      |      |      |      |
| Defense Payrolls .....     | 10.5 | 0.18 | 0.68 | 0.43 | 2.24 | 0.79 | 1.50 | 0.94 | 0.71 | 1.66 | 0.94 |
| Defense Contracts .....    | 19.7 | 0.15 | 1.34 | 0.75 | 1.62 | 1.36 | 1.20 | 0.67 | 0.29 | 1.25 | 0.95 |
| Defense-related Act. ....  | *    | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Atomic Energy .....        | 1.3  | 0.80 | 2.23 | 0.80 | 1.33 | 0.96 | 0.89 | 1.59 | 0.93 | 1.44 | 1.22 |
| Space Res. & Tech. ....    | 2.0  | 0.04 | 1.27 | 0.12 | 2.80 | 1.86 | 0.55 | 1.18 | 0.06 | 1.50 | 0.38 |
| Health Research .....      | 0.7  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Higher & Science Ed. ...   | 0.7  | 0.67 | 0.85 | 0.96 | 1.10 | 1.00 | 1.16 | 1.35 | 0.62 | 1.66 | 1.14 |
| Other                      |      |      |      |      |      |      |      |      |      |      |      |
| Internatl. Relations ..... | 1.0  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Postal Service .....       | 3.4  | 0.61 | 1.87 | 1.73 | 0.69 | 1.35 | 0.88 | 0.73 | 0.72 | 1.88 | 1.13 |
| Health Fac. Constr. ....   | 0.3  | 0.60 | 1.15 | 1.33 | 0.76 | 0.94 | 1.17 | 1.09 | 0.83 | 1.34 | 1.04 |
| Health Serv. & Care .....  | 1.7  | 0.60 | 1.62 | 1.69 | 0.89 | 1.26 | 0.96 | 0.85 | 0.69 | 2.00 | 1.16 |
| Misc. Health Act. ....     | 3.6  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Retirement Payments ...    | 16.4 | 0.83 | 1.14 | 1.19 | 0.88 | 1.03 | 1.01 | 1.01 | 0.92 | 1.30 | 1.07 |
| Other Income Security      |      |      |      |      |      |      |      |      |      |      |      |
| Payments .....             | 3.2  | 1.50 | 1.73 | 2.19 | 0.78 | 1.14 | 0.78 | 0.88 | 1.13 | 1.60 | 1.40 |
| Soc. & Ind. Services ..... | 1.3  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Veterans Programs .....    | 3.3  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Law Enforcement &          |      |      |      |      |      |      |      |      |      |      |      |
| Justice .....              | 0.2  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Debt Service .....         | 6.8  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| Ed. Loans .....            | 0.1  | 0.71 | 0.53 | 0.65 | 0.54 | 1.22 | 0.92 | 1.14 | 0.67 | 2.99 | 1.38 |
| Total Fed. Exp. ....       | 100  | 0.61 | 1.31 | 1.15 | 1.24 | 1.14 | 1.03 | 0.90 | 0.76 | 1.50 | 1.03 |

\* Less than 0.1%.

a FY 1969.

b All of these counties were declining in population.

Note: "... " indicates data are not available.

Concentration ratio is the ratio of the share of expenditures to the population in that type of area. May be interpreted as a per capita relative share.

Source: Adapted from U.S. Office of Management and Budget, Evaluation Division, "Location Analysis of Federal Expenditures in Fiscal Year 1969," September 1970. (Mimeo-graphed.)



### SEMI-DIRECT POPULATION DISTRIBUTION POLICIES

Half-way between direct programs and policies and implicit ones, there lie a number of others which might be called semi-direct, which have territorial issues of development among their objectives. These include the Federal Highway Program, which is generally recognized as having had a strong impact. Yet this program may now be essentially completed or mature: The system of metropolitan areas is by now well linked by roads. Further, as with many programs except the most direct ones, the effects of highways have often run counter to expectations. In many depressed areas, better connections have proved that roads lead out as well as in: Efficient transportation opens depressed area markets to competition from more efficient producers and encourages producers in the depressed area to ship their primary products out of the area without local elaboration. Similarly, within rural areas the program appears to have had the effect of thinning out the number of market towns in the areas served by extending the effective radius of movement.

Other semi-direct programs and policies include those dealing with the constellation of conservation, environment, and national recreation; they also include the location of the coming generation of international airports, which will be quite few; the location and mode of powerplants, which will be a great many; and water-oriented developments, such as new sewage disposal technologies, the control of precipitation through cloud-seeding, the shifting of waters regionally and internationally through massive projects, as well as the more traditional navigation, irrigation, and flood-control projects. The territorial intent of such projects was evident at the opening ceremonies of the Arkansas River Navigation project in 1971: President Nixon said of the newly maritime states of Oklahoma and Arkansas, "This region can become a new magnet for people seeking the good life, so that we begin to see a reversal of decades-long migration from rural America to urban America."

Such proposed programs as revenue sharing are also semi-direct programs, which will affect the level of public services, the tax-load, and the degree of local self-determination throughout the country; the Federal procurement policies discussed above (of which the 1971 Lockheed controversy is an extreme example); and the welfare laws (as designed by Congress and the states, and interpreted by the courts). Of these last, the trends towards equalization are viewed as potentially powerful retardants of migration towards the cities and even as

reversers of these trends, while the striking of residency requirements by the courts are viewed as accelerators. Nor are these all of the effects anticipated from welfare reform: They are viewed as complements of the minimum wage laws, affecting marginal industries in marginal locations, and as providers of a degree of economic independence to the poor, with consequences as to their willingness to take political and economic risks.

Semi-direct policies also include agricultural policies, many of which have traditionally been based as much on life-style considerations, such as preserving the family farm, as on economic considerations. A recent analysis argues that beginning in the 1940's, price supports and acreage limitations, together with agricultural loan policies, encouraged a substitution of capital for labor in cotton, the consolidation of holdings, and the shift from sharecroppers to wage-workers.<sup>5</sup> This produced a surplus of agricultural laborers which increased dramatically the migration to urban centers. Thus, the racial and poverty problems of urban centers today can be considered the consequence of the exigencies on labor of World War II and of the modernization of agriculture in the South, abetted by agricultural policies in the post-war decades. If this analysis is correct, and the overt intent of the policies was the maintenance of an agricultural population, the conclusion must be that the policy had the result opposite (if race and class is neglected) from that intended.

### OTHER POLICIES AFFECTING POPULATION DISTRIBUTION

In all of these instances of what we have called semi-direct policies, the regional effects are considered, even if they are not paramount; and the complicated interrelations of the social system often produce regional consequences unforeseen or markedly different from those intended. Although the difference is one of degree, there are also policies whose territorial consequences are less frequently taken into account but which have, at least potentially, strong territorial impacts. These territorial impacts are called implicit policies. The analysis of their effects in this paper is, of necessity, fairly speculative given our present level of understanding; the instances presented below should be regarded as illustrative.

One of the central preoccupations of national domestic policy is the cyclical control of the economy. In recent years, there have been shifts in emphases between the use of monetary policy and fiscal policy, and within fiscal policy between an emphasis on tax

inducements to investment and on direct employment in public works. Whatever the merits of these policies as counter-cyclical measures, their regional consequences may be quite different. Tax credits for business investment favor localities whose economy is based on capital-intensive activities with a capacity for rapid adoption of new capital. Since it appears that smaller cities specialize in well-established, traditional industries, whose technology has developed capital-intensive forms along well-established lines, it might be expected that such localities would especially benefit. Similarly, fast growing areas, which need capital for new enterprises, would be expected to benefit. On the other hand, an increase in the supply of money and a lowering of the discount rate (or the reverse) under monetary policy would affect most sharply fast growing regions, both for public and private investment, including consumer investment in the form of housing and durables. Swings in the interest rates of municipal bonds might affect local tax rates and borrowing capacity in such regions for several decades, thus affecting their attractiveness in terms of levels of public service, their ability to invest in physical and social infrastructure, and so forth. Comparably, fiscal measures based on lowering taxes will have different effects from public works ones, since the ability to mount rapidly substantial programs will vary greatly from region to region, as will the usefulness (other than counter-cyclical) of the projects. General tax reduction, including income tax reduction, would presumably favor large, high-income urban areas with stable rates of growth.

The general field of taxation has many regional implications. Such matters as depreciation allowances for oil are obviously regionally concentrated. Other provisions, such as the treatment of losses in farm operation, affect sharply the economies of marginal farming regions near urban centers. Provisions for accelerated depreciation of capital encourage capital investment, and shift resources to regions whose economic activities specialize in capital-intensive activities—typically provincial centers. By contrast, the treatment of research and development expenses as ordinary expenses and of most of the gains resulting from these as ordinary income, sets regions which specialize in innovation at a comparative disadvantage; at a national scale such treatment encourages the expansion of existing technologies rather than development of new ones, with consequent regional effects. On the whole, these provisions appear to advance interregional equity by favoring provincial regions, but to be regressive by favoring existing technologies over new ones.

In addition, tax laws affect territorial development by their treatment of entertainment and the like as business expenses to a degree possibly not realized. This has given rise to the dual society of the expense account and the deductible evening out. Although no studies appear to exist on the subject, this practice clearly underlies much of the glamour and the economy of certain metropolitan centers. Manhattan would not be the same without it, nor any other major metropolitan center. It may be presumed that, were these laws tightened, our socio-economic landscape would be grayer but more dispersed.

Although this discussion is aimed at large-scale regional effects, several intrametropolitan effects must be mentioned. The tax break to homeowners, through the deductibility of mortgage interest and property taxes, has been recently estimated by the Treasury Department as amounting to \$5.7 billion annually and undoubtedly influences the tendency toward single-family homes and low densities. The attribution of local school costs to local property taxes (and to some degree the costs of welfare) also shapes metropolitan areas by encouraging mercantilistic policies by local taxing districts to exclude the poor and the fertile young and to capture industrial and commercial activities. Within the central cities, the provisions of our tax laws, which permit indefinite multiple depreciations of old buildings and capital gains on their resale, undoubtedly affect the density of population, the maintenance of stock and its abandonment, and the rent levels which people must pay. The structure of our metropolitan areas reflects our tax laws much as English architecture reflected the window tax of the middle ages by having fewer and bigger windows.

It seems highly probable at this writing that we shall soon have some form of polluter's taxes, where an activity is taxed according to its contribution to the problem. It will make a great deal of difference to the distribution of regional development whether these taxes are at uniform rates at all locations, or whether they are graduated according to the severity of the local pollution problem. If uniform rates are applied, a firm has a choice of paying these taxes to compensate the public for the negative externalities it imposes, or of incurring additional capital and operating costs to abate its emissions. But if the rates vary with the severity of the local problem, firms have a third option: to move to locations where the environment is less burdened and better able to regenerate itself. Many firms would relocate, especially manufacturing ones since they are heavier polluters than service sector enterprises and are less

linked to metropolitan economies. The net result would undoubtedly be considerable dispersion of economic activity, much of it towards currently depressed areas. Earlier in this paper it was remarked that population dispersal policies are slow and inefficient instruments against pollution. But, conversely, it would appear that anti-pollution measures might be fairly effective population-dispersal instruments. Yet, although this result would increase interregional equity, its costs might be borne to a large degree by the working classes in the metropolitan areas, who would lose many jobs.

Local differentials in pollution taxes might be expected to have further consequences. Where population densities are low, automobile emissions are not a serious problem, so that using pollution taxes as disincentives against cars makes no great sense. Should taxes on automobile emissions be high in the large urban areas and low in less dense ones, it would confer a relative advantage to these latter. Within metropolitan areas, heavy taxes would be a strong incentive to switch from the automobile to other forms of transportation; this would modify the grain and structure of urban settlements.

Examples of implicit policies can go on and on, since the point being made is that virtually all Federal policies have territorial consequences. Or, put another way, the territorial distribution of our population and our economy is the projection on the geographic plane of our socio-economic system. A few more instances will be given in the following pages, without trying to be comprehensive, and some suggestions will be offered in conclusion.

The United States has been engaged in major wars in Asia for the greater part of three decades. That this has had a major effect on the development of the West Coast is so obvious it is often overlooked. The point need not be labored, except to stress its magnitude: With the possible exception of the continuing agricultural revolution, it is probably the most significant factor that has affected the distribution of people and activities for the past third of a century. Similarly, the decision to put a man on the moon has had profound consequences for Houston, for Florida, for Cambridge and Palo Alto, and for many other localities. The effect of this can be glimpsed in Table 1, where defense payrolls, defense contracts, and space research and technology are so large and so concentrated in the fastest growing counties that they alone account for the fact that federal expenditures are most concentrated in the fastest growing counties. Symmetrically, the winding down of the Asian war and the de-emphasis of aerospace research account for the

pockets of depression currently resulting in many of these areas.

Yet Federal expenditures in certain localities do not always result in straightforward local benefits. It will be noted in Table 1 that Federal expenditures for water transportation are concentrated in the central cities of metropolitan areas to a degree unmatched by other expenditures in the table. A great deal of this has to do with the modernization of ports for the new technology of containerization. Yet this is a mixed blessing for those central cities. To be sure, construction expenditures add temporarily to the income of the area. But in the long run, the effects may be quite different. Since the goods are boxed, a great deal less labor is needed per ton or per dollar. Thus, direct employment for port activities will fall. The opportunities for break of bulk and additional transformation for trans-shipment activities are greatly reduced at port cities. These activities tend to move inland, together with the preparation for shipment, closer to the sources or destinations of goods. On the whole, the modernization of shipping takes away from the importance of the ports and gives added importance to inland productive locations. As a further complication, the present economics of this technology result in the by-passing of secondary ports and the concentration of shipping through primary ones.

Comparable to this effect is that of the tendency toward larger and faster passenger aircraft. This tendency arises not only from the logic of the aircraft and airline industries in the private sector, but also from a complex interplay of governmental defense and procurement policies. The logic of faster and larger aircraft is heavier volumes, longer runs, and fewer stops. Consequently, the direction of this effect is concentration towards the principal urban areas, especially for those activities, such as management, marketing, and technology, that require interpersonal contact.

The effects of the developments in water and air transportation will be comparable in the magnitude, if not necessarily in the direction of their effects, to those of the massive highway developments of the last two decades, which are generally thought to have been of profound significance for the distribution of economic activity and the movement of population.

Transportation and communication are also strongly affected by regulatory commissions, which to a very large extent set up the rules of the game. Their decisions as to pricing and the conditions of supply by the industries they supervise strongly influence regional development. (An obvious instance of this is the matter of fuel oil for New England.) These commissions seek to be even-handed and judicial, and are generally unmindful

of the territorial or regional development effects of their decision.<sup>6</sup> Clearly, decisions and practices of the commissions that deal with interstate commerce and transportation have been most important for the distribution of manufacturing. It may be expected that the decisions of the Federal Communications Commission will gain in importance for the distribution of economic activities.

Manufacturing has become a relatively stable sector in terms of national employment; most national and regional economic growth is based on the expansion of the service sector. The service sector, in such areas as insurance, inventory, management, and government, is not only growing but becoming more like manufacturing. By standardizing information and routinizing its handling, it can ship data and ideas as if they were things, and employ semi-skilled labor to process them. The locations of such activities have come to resemble more closely those of manufacturing, leaving the large urban centers in search of other advantages. The extent to which they will do so will depend to a large extent on how the new technologies of data transmission are organized and on the pricing policies instituted, much as the distribution of physical production has depended on the structure of the transportation matrix. In some ways, the almost exclusive concern of regional development policy with the location of manufacturing is fighting the last war. The distribution of the service sector looms as a larger factor.

The support of higher education and research has had dramatic effects for regional development in several areas, of which Cambridge and Palo Alto are perhaps prototypical. The recognition of the catalytic effects of intellectual centers for regional development was one of the reasons for the adoption some years ago of policy of dispersing such support to many "centers of excellence." It would appear that this essentially regional policy has resulted in an "implicit policy" for the production of scientific personnel. The current glut of scientific manpower is being made harder to manage by the coming to maturity of Ph.D. programs in many second-rank universities, which had been supported under the "centers of excellence" policies. At the same time, the technology-oriented regional development in Cambridge and Palo Alto has resulted in localized depressions compounded from the national, general economic slowdown and the concentration of this slowdown within the technology sector. Yet, in the long run, it is obvious that the location of the knowledge-producing industry (which is largely supported from Federal funds) is an important determinant of the location of some of the most dynamic sectors of the national economy.

One report mentioned earlier concluded that "investment in human resources... has little direct and demonstrable economic impact,"<sup>7</sup> and observed that "the Departments of Labor and Health, Education and Welfare are not greatly concerned about the economic development of localities in these programs. They are exclusively client-oriented. Hence, the funds go where client needs and demands are."<sup>8</sup> Yet education has a well-known effect on the population of distressed areas: The higher the level of education of an individual, the higher are the chances that he will leave the area. Thus, education programs do not have a significant impact (in terms of economic growth) on the development of a distressed area. Yet they do have an effect by thinning the population, by providing those who leave with an opportunity for social mobility through geographic mobility, and possibly by reducing the competition for opportunities among the remaining population. These are strong territorial effects, although education is not a prime instrument for development of poor areas.

Finally, national immigration policy strongly affects particular locations. Net civilian immigration accounts for one-fifth of national population growth. Miami will always bear the imprint of the Cuban inflow of the past decade. San Francisco has strongly felt, in its Oriental population, the consequences of the liberalization of the ethnic provisions of the law. Southwestern agriculture is strongly affected by the treatment of permanent and temporary Mexican migration. Although the rate and composition in migration has been as strong a force as any for molding the regional structure of this country, and continues to be a considerable one, it appears to be one of the least studied from the point of view of general national domestic policy.

## CONCLUSIONS AND RECOMMENDATIONS

We have argued that our society and economy are a highly interconnected system, of which the geographic distribution of population and economic activities is only one aspect. Direct policies to modify this geographic distribution have been generally ineffective and sometimes counterproductive because they have underestimated or misjudged the connections among elements of the system along dimensions other than the geographic. Conversely, many policies and public actions whose main thrust is not territorial turn out to have strong geographic consequences which are normally not intended. From the point of view of distribution, these are "implicit policies."

In the case of direct and semi-direct policies, it is possible to conceive of increasing coordination of

programs and clarification of objectives. Substantial efforts in this direction are being made, including the formation of the President's Domestic Council, the formation of Regional Commissions, the redesign of Federal administrative districts to increase the congruence among them, and efforts to make Federal agencies more cooperative with each other, with local agencies, and with the public.

But in the case of implicit policies, it is neither realistic nor desirable to force them into formal regional coordination. Their effects on geographic distribution are only one consideration for these, and usually not the main one. It would seem more appropriate to have a form of indicative planning that would stress information as to national geographic consequences of current or contemplated national policies and programs for these issues. Geographic distribution cannot be the unifying theme of all national policies. What seems possible is to improve the level of awareness by the public and policy makers of the geographic consequences of actions which are not primarily geographic in intent.

In practice, this function has been carried on largely by Congress, which is organized on the basis of geographic representation. Many of the instances of implicit policies which served as illustrations in the pages above were in fact shaped by geographic interests in the process of their formulation. Yet this has been too often done on the basis of shifting political alliances, short-sighted perceptions of interest, and not uncommonly with an incorrect understanding of consequences.

Title VII of the National Housing Act directs the President, using the capacity of the Domestic Council to gather and analyze information, to present to Congress every other year beginning in 1972 a Report on Urban Growth. The language of the Act suggests that this report would focus primarily on direct policies. This seems appropriate in that the President, as Chief Executive, presides over the diverse direct programs and is responsible for their appropriateness, effectiveness, and coordination.

Yet we have noted the special relation of a territorially organized Congress to implicit territorial policies. It would appear useful to have an agency less directly linked with the executive branch which could report directly and frequently to the President, to Congress, and to the public on territorial problems and on the implicit territorial consequences of diverse national policies. This report would include analysis of proposed policy and pending legislation of regulatory commission decisions and of administrative practices in the implementation of existing legislation. It would also include longer-range studies of the consequences of

long-standing policies and of emerging problems and surveys of public preferences and perceptions of problems, including the canvassing of preferences and opinions of diverse citizen groups and public agencies. This agency would emphasize implicit policies, thus complementing the President's Report on Urban Growth.

Although the establishment of such an agency might take many forms, from a division of the Office of Management and Budget to an expansion of the role of the Urban Institute, it is worth noting that under Section 703(c) of the Housing Act, the President is empowered to establish an advisory board, composed of scholars and Federal, state, and local officials, to assist in the preparation of the Report on Urban Growth and any supplementary reports. Such an advisory board, if established on a permanent basis and supported by a permanent staff, might be able to perform the indicative functions of analysis of problems and of the consequences of policy. It is clearly important that the funding of this agency be general and long range, since its purposes are evaluative and its functions may at times be those of criticism. Obviously, such an agency could not maintain its integrity if dependent on grants and contracts which thrust it into a client relationship.

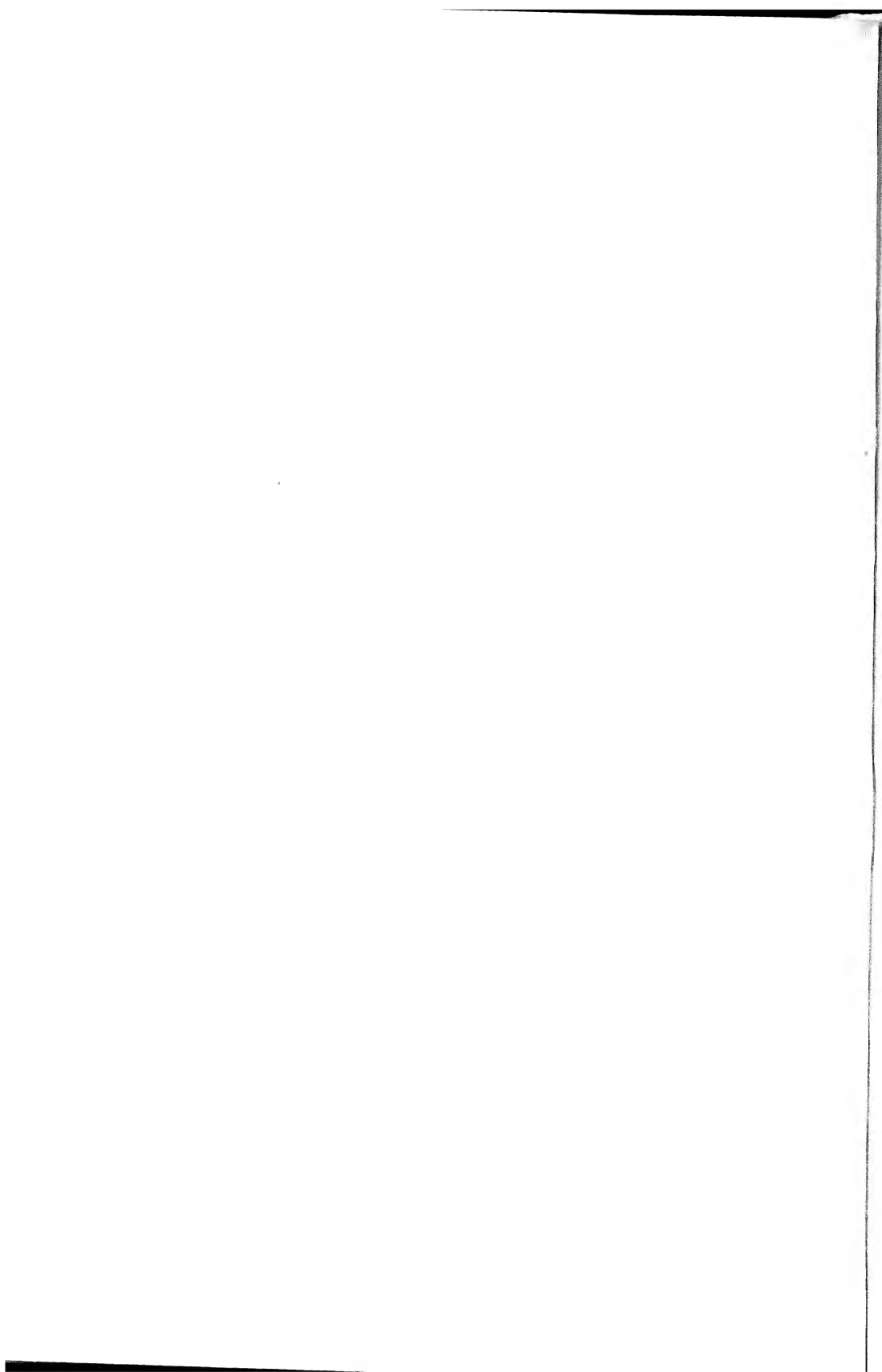
The beginning of this paper stressed that our present understanding is poor with respect to ongoing processes, to the definition of problems and of policy objectives, and to the consequences of diverse public actions and policies. Under these circumstances, mechanisms to facilitate the process of social learning (rather than master plans) are urgently needed. The proposed agency would aid this learning process, both directly and by stimulating a new awareness of territorial issues throughout government and society.

## REFERENCES

1. U.S. Advisory Commission in Intergovernmental Relations, *Urban and Rural America: Policies for Future Growth*, 1968, p. 5.
2. I am referring here to large-scale territorial policy in the case of poverty and discrimination. At the smaller scale of intrametropolitan policy, school and residential integration for access to jobs are extremely important territorial issues.
3. For other nations in other circumstances other objectives are relevant. Among these are the occupation of territories near disputed borders (an objective that played an important part earlier in U.S. history from upstate New York to southern Texas), the reduction of economic colonialist dependence, the creation of national identity, the assimilation or preservation of sub-national ethnic groups, the territorial consistency-check on sectorally-specified national plans, and, in complex ways, the integration of implementation and planning in directed economies.

4. U.S. Dept. of Commerce, Economic Development Administration, *Federal Activities Affecting Location of Economic Development*, prepared by the Center for Political Research, Nov. 1970, Vols. I and II.
5. C. F. Hale, "Impact of Federal Policy and Technological

- Change on Regional and Urban Planning Problems," *Land Economics*, February 1971.
6. U.S. Dept. of Commerce, *op. cit.*, Vol. I, p. 5.
7. *Ibid.*, p. 4.
8. *Ibid.*, p. 17.



---

**Part IV Case Studies of Regional Planning**



1

2

3

4

5

## Introductory Note

The last decade has seen a proliferation of case studies of regional policy and planning. We thus begin to be able to examine theories of regional development and urbanization in practice. But most of the studies can only describe the historical origins of regional planning, its formal mechanics, the policies and instruments adopted, and the political dynamics that underlie them. It is still too early to judge the lasting effectiveness of these policies.

The study of Turkey by İlhan Tekeli (chapter 30) is distinguished by its historical perspective and makes strikingly clear that in Turkey there has been no simple linear path of evolution in spatial organization. The process has been one of complex cyclical patterns in the centralization/decentralization of power and of spatial integration/fragmentation, reflecting historical changes in technology, social institutions, and external relations. By contrast, Elisabeth Altman and Betsey Rosenbaum show how the antiurban ideology of Zionism, combined with the need to occupy the national territory for defense purposes, resulted in a policy of dispersed settlement (chapter 31). And Robert Fried shows how in Italy a series of impending sociopolitical crises molded the character of regional planning in that country, so that Italy's spatial policies were formed in the conflict and competition among regional planning agencies and their interests (chapter 32).

The chapters concerned with Spain and France deal with regional planning as an extension of comprehensive national and sectoral planning (chapters 33 and 34). H. W. Richardson discusses Spain's efficiency-minded policy of "growth poles" initiated in 1964 (along with deconcentration policies for Madrid), while George Ross and Stephen Cohen trace the political history of postwar regional planning in France. Like their Spanish colleagues, French planners made use of the growth pole concept, but unlike them they were more concerned with the territorial organization of power and the need to stimulate economic and cultural life in *le desert français*, the periphery to the all-powerful Parisian core. The analysis is of particular interest for its exploration of the interaction of technical planning and political processes.

The final two contributions present an interesting contrast between the centrally managed economy of the Czechoslovak Socialist Republic (chapter 35) and the market economy of the United States (chapter 36). D. Schejbal and O. Žurek describe the elaborate procedures of coordination and target setting, whose purpose is "proportionality in the development of the national economy" (akin to the American usage "balanced growth") and continuously rising living standards. Although the procedures for regional planning

vary importantly among the socialist countries, this Czech example is representative of the principal style. We have not found a presentation with more information as to the politics, conflicts, or social factors at play and more insight into the frictionless machine portrayed here.

By contrast, William Alonso's critical overview of subnational planning in the United States raises social and political problems (chapter 36). It shows that American regional planning by the national government has been primarily reactive, almost exclusively concerned in recent years with economically distressed regions left behind by the market forces of development.

Readers who are familiar with current developments in regional planning in these countries will be well aware that all of these case studies carry the story forward only to a certain point, and that much has changed in several cases. This is unavoidable, but perhaps it is less important than it might seem at first. These examples serve to give a flavor of the practice of regional planning, and we hope that they provide a background by which the reader can see more recent developments in perspective. No universal success formula arises, but the studies give insight into the varieties of regional planning experience and the multiplicity of influences that act upon it.

---

## 30 Evolution of Spatial Organization in the Ottoman Empire and Turkish Republic

Ilhan Tekeli

**D**URING the sixteenth century, the boundaries of the Ottoman Empire extended over the Near East, North Africa, and the Balkans. Because of its geographic location, the Ottoman Empire came into early contact with the changes taking place in the West, and, at the same time, it was the last feudal empire of Europe when industrialization began; and because it could not make the necessary adaptation to industrialization, it passed to a state of rapid decay after the eighteenth century.

Today there are about two dozen nations living within the former boundaries of the Empire which have passed through different stages of development. This chapter deals with the changes which have occurred in the spatial organization of one of these nations, Turkey, from the period of the Ottoman Empire to the present. In the analysis, the Ottoman social structure and spatial organization of the sixteenth century are taken as the reference point. Although there were important regional disparities, this study may offer a historical parallel for the experiences of other countries.

Our task will be to analyze how the settlement structure of the Ottoman Empire changed and which factors influenced this process. In historical perspective, we shall see that a change in the Ottoman settlement structure coincided with the transition to the "semi-colonial"<sup>1</sup> period during the nineteenth century. Later, the young Turkish Republic implemented deliberate settlement and regional development policies because of the unbalanced settlement structure inherited from the Ottoman Empire and an increasing number of ethnic problems. Although the spatial policies of the Turkish Republic were born as a reaction to the typical foreign relations of the Ottoman Empire and did show certain fluctuations in time,

<sup>1</sup> V. I. Lenin, *Imperialism: The Highest Stage of Capitalism* (New York, International Publishers, 1953), p. 10.

they remained quite stable on the whole. The essentials of the spatial policy laid down in the last two Five-Year Development Plans may be traced back to the beginning of the Republic.

#### SPATIAL ORGANIZATION OF THE OTTOMAN EMPIRE DURING THE SIXTEENTH CENTURY

The sixteenth century is the period when the Empire reached the limits of its expansion. The settlement organization was quite balanced and integrated within the central feudal order of the Ottoman society, bounded, of course, by the technological limitations of that period.

Land and labor were the scarce factors of production in a production pattern which was limited by the level of technology. The increase of the controlled surplus in such a pattern depended on (a) expansion in size of the cultivated land and controlled labor, and (b) a more rational use of land and labor, both of which are dependent on organization.<sup>2</sup>

In a feudal social system land and the size of the controlled group are expanded through conquest. As conquest itself is achieved through organization, so too is the control of the surplus of invaded areas and their integration into the existing system. An increase in organization and an increase in the size of the controlled groups led to a greater need and opportunity for urbanization. Therefore, a central authority was necessary in order to defend the position of the Empire on the main East-West road (the Silk Road), to achieve continuous expansion through conquests, and to govern a quite large urban population.

Because of the relatively stable level in the international technology of production, the establishment of central authority and supremacy depended on rationality in organization. This rationality was achieved through land tenure and military institutions<sup>3</sup> which developed as an interrelated harmonious system. A system of military fiefs enabled the Empire to expand its territories and the size of the controlled labor group through conquests, without the burden of supporting a substantial central army and with a minimum loss of the agricultural labor force. The military fiefs depended on the *miri* land system (domain land belonging to the sovereign power). Domain land and the state's ability to control the spatial distribution of labor through regulations, such as the *surgun* (official transfer of populations from one area to another) and those forbidding the *raaya* (subjects, in the present context, "peasants") to leave the land, made the rational use of land and labor possible. The central authority could thus maximize the surplus value it controlled with a control system.<sup>4</sup>

The allocation of power in an empire, such as the Ottoman Empire, which aims at increasing its power—i.e., increasing the controlled surplus through conquests—is

<sup>2</sup> Mübeccel Kıray, "Toplum Yapısındaki Temel Değişimlerin Tarihsel Perspektifi, Bugünkü ve Yarınki Türk Toplumunu Yapısı" [Basic changes in social structure in historical perspective—Turkish social structure today and tomorrow] in *Mimarlık Semineri* (T.M.M.P.B. Mimarlar Odası), 1969.

<sup>3</sup> İsmail Cem, *Türkiye'de Geri Kalmışlığın Tarihi* [History of Underdevelopment in Turkey] (İstanbul: Cem Yayınevi, 1970), pp. 53-55.

<sup>4</sup> The question of whether or not the local control elements were hereditary is not important for the rationality of the system. Adequacy of military forces and rational use of land and labor are important.

best represented by a "center-front" model.<sup>5</sup> The "front" grows constantly and is self-perpetuating. In a system like this, there are important spatial problems at the "front." Not only were the fighting forces of the Empire concentrated here, but also a new mixture of cultures was formed by the planned transfer of population through exile. These societies were open to integration, very heterogeneous and creative.<sup>6</sup> A hierarchy of control must be the basis for a central control system for an empire with a "front" like this, and a hierarchy as such must coincide with the hierarchy of spatial organization.

In sixteenth-century Ottoman society one finds a hierarchy of settlements consisting of a capital, regional centers, market cities-villages, and nomad groups. Such a hierarchy depends on the functional differentiation of cities and the relation between such cities and their hinterlands and other cities as well.

According to the economic historian Ömer Lütfi Barkan,<sup>7</sup> the population of Istanbul, which was 97,956 in 1478, grew to 400,000 in the period 1520-1535, thus raising Istanbul to the rank of largest city in Europe. Although the transportation and production technology of that period made the production of only a limited agricultural surplus possible, nevertheless, all the needs of a big city even as big as Istanbul could be met<sup>8</sup> by the excellent organization of the social system. The efficient organization of the transfer of surplus to the capital, Istanbul, and the integrated state of the Empire were the main factors contributing to the unusual growth of that city.

The existence of large groups settled<sup>9</sup> in Istanbul makes it apparent that despite the limited technology of the time there were no bottlenecks in obtaining a food supply for this large city. In later periods (around the seventeenth and eighteenth centuries), when food shortages began, precautions were taken to slow down the growth of the city.<sup>10</sup>

Istanbul was quite a differentiated city as a result of its concentration<sup>11</sup> of administrative, military, commercial and industrial as well as religious and social

<sup>5</sup> This model is different from the "center-periphery" models used in development theories. In a "center-periphery" model the integration is between a growing center and a periphery which in general is declining, whereas in a "center-front" model there is a growing, developing front.

<sup>6</sup> Paul Wittek, *The Rise of the Ottoman Empire* (London: Luzac and Co., Ltd., 1963), pp. 33-34.

<sup>7</sup> Ö. L. Barkan, "Essai sur les données statistique des registres de recensement dans l'Empire Ottoman aux XVe et XVIe siècles," *Journal of Economic and Social History of the Orient* 1 (August 1957): 20-21, 27.

<sup>8</sup> At this point one must remember that it was possible to transport goods to Istanbul via waterways.

<sup>9</sup> Among the groups settled in Istanbul were 40,000 Armenians coming from Crimea and Kefe (after the conquests of Crimea in 1475) and Bulgarians who settled in the areas between Yedikule and Topkapı and Büyükdere after the conquest of Serbia in 1520. Muslim, Arab, and Jewish groups settled in the Balat, Hasköy, and Ortaköy districts after the fall of Granada in 1492. See Osman Ergin, *Türkiye'de Şehirciliğin Tarihi* [History of Urbanism in Turkey] (Istanbul: İstanbul Üniversitesi Hukuk Fakültesi İktisat ve İçtimaiyat Enstitüsü publication no. 3, 1936), pp. 96-97.

<sup>10</sup> Sabri F. Ülgener, *Tarihte Darlık Buhranları ve İktisadî Muvazenesizlik Meselesi* [Crisis of scarcity in history and the question of economic unbalance] (Istanbul: İstanbul Üniversitesi İktisat Fakültesi, 1951).

<sup>11</sup> Some data from the seventeenth century give an idea about the size of Istanbul in the sixteenth century. When Evliya Çelebi, the well-known traveller and historian, described a procession in Istanbul, he counted about 1,100 different occupations. The people involved in these occupations were workers, merchants, craftsmen, and others who were employed by the service sector (*Evliya Çelebi Seyahatnamesi*, edited and translated by Zuhuri Danişman, vol. 2, p. 206). According to R. Mantran, in the sixteenth century more than 10,000 laborers were working in 28 public workshops, whereas in the 25,000 private sector workshops 80,000 laborers, including the employers, were working. It was necessary to allocate the agricultural surplus of Rumania, the Danube coastal regions, and most of the Rumelian Sancaks to Istanbul. See İsmail Cem, op. cit., p. 66. Cem himself is relying on information in Robert Mantran, *Istanbul dans la seconde moitié du XVIIe siècle* (Paris: Librairie Maisonneuve, 1962), p. 188.

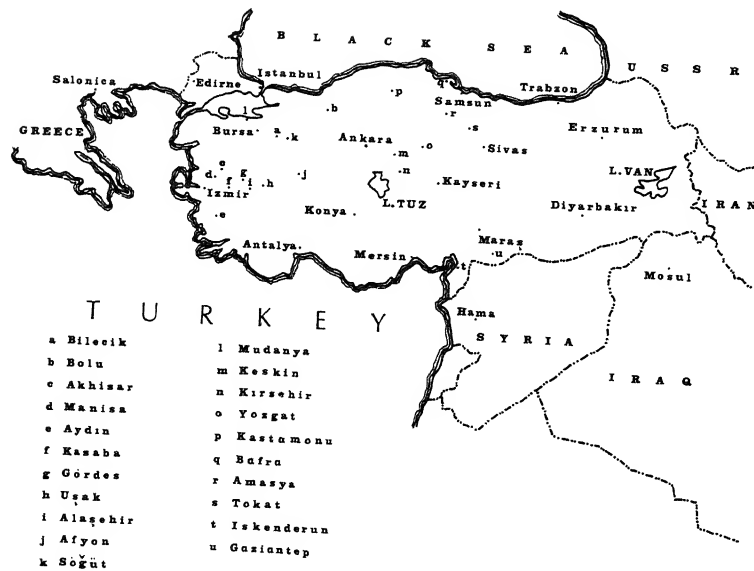


Fig. 9. Sketch Map of Turkey.

institutions. Although both capital and frontiers gained military and administrative power, economic power and local control and organization still depended on the existence of regional centers, which were quite evenly distributed over the territory. During the sixteenth century, the population of such centers varied between 20,000-40,000 inhabitants and were located at intersections of important trade routes or at breaking points. For example, Sivas (20,000), a regional center, was located at the junction of the north-south route (Crimea, Syria, Egypt) and the east-west route (Antalya, Konya, Kayseri, Sivas, Erzurum, Tabriz); it was famous for wool and cotton textiles and especially for mohair. About 15 different types of small industrial workshops and specialized commercial shops were located in the city.<sup>12</sup>

The Bursa-Bilecik region was known for silk weaving. Bursa, with a population of 40,000,<sup>13</sup> was not only a silk market but also a warehouse<sup>14</sup> for buying and transporting the cotton textiles of Western Anatolia to Istanbul, Rumelia and Southern Russia and the mohair of Ankara and Kastamonu to Europe through

<sup>12</sup> Osman Turan, "Selçuklular Zamanında Sivas Şehri" [Sivas under the Seljuqs], Ankara Üniversitesi Dil ve Tarih Coğrafya Fakültesi Dergisi, vol. 9, no. 4 (1951), pp. 449-451, 456-457.

<sup>13</sup> H. İnalcık, "Osmanlı İmparatorluğu'nun Kuruluş ve İnkişafı Devrinde Türkiye'nin İktisadî Vaziyeti" [Economic conditions in Turkey at the time of the establishment and expansion of the Ottoman Empire], Belleten, vol. 15, no. 60 (October 1951), p. 607.

<sup>14</sup> H. İnalcık, "15. Asır Türkiye İktisadî ve İhtimal Tarihî Kaynakları" [Sources of the economic and social history of fifteenth-century Turkey], İstanbul Üniversitesi İktisat Fakültesi Mecmuası, vol. 15 (October 1953), pp. 51-55. See also H. İnalcık, "Bursa and the Commerce of the Levant," Journal of the Economic and Social History of the Orient, vol. 3, pt. 2 (August 1960), pp. 131-147.

Western middlemen. In this case one can clearly follow the correspondence of the location on the junction of north-south and east-west trade routes to the development of regional center functions.

Diyarbakir, located on the road to Trabzon-Mosul-Baghdad, had a population of 12,000<sup>15</sup> in 1518, and specialized in textiles and leather.

In 1520, Edirne's population reached 22,335.<sup>16</sup> Edirne was not specialized in any industry; yet, its population increased because of trade and the city's function as a second capital (in time of war).<sup>17</sup>

Geographic specialization in production takes place not only in industrial societies but also in feudal ones. The Ankara and Kastamonu areas were specialized in mohair weaving; the Konya-Afyon area in mat weaving; and the Usak-Gördes area in carpet weaving.<sup>18</sup> The Bursa-Bilecik area was noted for its raw silk and for weaving; Mesopotamia for its oil lamps for mosques and for its glass-ware; Manisa-Akhisar-Alasehir regions for tanning and leather products;<sup>19</sup> Maras for its wrought iron; Damascus for its iron works, sword manufacturing, and enamel; and Gaziantep for its footwear. Ankara was also specialized in the manufacture of caravan tools. Thus, in addition to performing transit functions on important routes, these cities, because of the geographic specialization, created commodity flows and performed commercial functions.

Such regional centers, besides specializing in one or more types of production at the interregional level, also produced goods for the consumption of their immediate hinterlands. For example, while there were 25 different guilds in Manisa,<sup>20</sup> the city showed geographic specialization in only one or two of these. These villages, besides their agricultural activities, also specialized in those activities which the central city was noted for. Thus the settlement areas functioned as integrated<sup>21</sup> production entities.<sup>22</sup> Geographic specialization also produced a vertical integration stemming from the local raw material: for instance, animal husbandry, tanning and leather manufacturing, footwear manufacturing; or raising mulberries, raw silk production, silk weaving; or Angora goat breeding, mohair, textiles.

<sup>15</sup> Nejat Göğünc, "Onaltıncı Yüzyılın İlk Yansında Diyarbakir" [Diyarbakir in the first half of the sixteenth century] *Belgelerle Türk Tarihi Dergisi*, no. 7 (April 1968).

<sup>16</sup> Halil Sahillioğlu, "XVIII Yüzyılda Edirne'nin Ticari İmkânları" [Prospects for commerce in Edirne in the eighteenth century], *Belgelerle Türk Tarihi Dergisi*, no. 13 (October 1968).

<sup>17</sup> Hasan Reşit Tankut, *Köylerimiz* [Our Villages] (n.p. 1939), pp. 12-13. Taken from Doğan Avcioğlu, *Türkiye'nin Düzeni* [The system of Turkey] (Anakar: Bilgi Yayınevi, 1968), p. 12.

<sup>18</sup> The Ottomans transferred their capital from one city to another according to the direction of growth of the empire toward Europe. Söğüt, Bursa, Edirne, and Istanbul became capital cities in that order.

<sup>19</sup> İbrahim Gökçen, *Manisa'da XV ve XVIII Yüzyılda Deri Sanatları Tarihi* [The History of the arts of hide working in Manisa in the fifteenth and eighteenth centuries] (C. H. P. Manisa Helkevi publication, 1945).

<sup>20</sup> Çağatay Uluçay, *17. Asırda Manisa'da Ziraat, Ticaret ve Esnaf Teşkilatı* [The organization of agriculture, commerce, and crafts in Manisa in the seventeenth century] (1942).

<sup>21</sup> The Polish traveller Simeon, who visited the Ottoman Empire during 1608-1619, tells about a city called Yenipazar near Üsküp in which he observed this integration between the village and the city in the manufacture of locks. Hrand. D. Andreasyan, *Polonyalı Simeon'un Seyahatnamesi* [The travel accounts of Simeon the Pole] (Istanbul: İstanbul Üniversitesi Edebiyat Fakültesi publication, 1964).

<sup>22</sup> Sencer Divitçioğlu, *Asya Üretim Tarzı ve Osmanlı Toplumu* [The Asiatic mode of production and Ottoman society] (Istanbul: İstanbul Üniversitesi İktisat Fakültesi publication, 1967), p. 50.



Regional centers were cultural and social centers as well. Educational, health, and social welfare services were supported by government-supervised religious trust funds. In the sixteenth century, the resources allocated for the support of such services consisted of 17% of the income of the Empire.<sup>23</sup> The distribution of these social service functions ran parallel to the hierarchy of settlements.<sup>24</sup>

This regional structure, which shows such geographic integration and specialization of production, engendered an urban hierarchy. So far we have differentiated two urban echelons: the capital and the regional centers. From documents of *nüzül*<sup>25</sup> (a kind of extraordinary tax levied for military campaigns) of the year 1590, it is possible to reconstruct two more hierarchical levels of settlements: those with a population of less than 10,000, such as Isparta (8,000),<sup>26</sup> and those with a population of less than 5,000. A fifth echelon consisted of villages integrated into this structure by means of a definite division of labor.

Since the state was the owner of the land and the peasant did not have the legal right to leave it, he had to cultivate the plot set aside for him and fulfill his social obligations. He was not free to choose which crops to grow or whether to grow anything at all. The type of crop to be raised in each region was decided by a central authority, and the product was allocated for the consumption of a certain city.

According to Ömer Lütfi Barkan, "the villages were responsible for growing the best barley or rice according to a definite plan, by share cropping. Many villages had to perform traditional duties inherited from their ancestors, such as producing alum, saltpeter, living in the area around caravanseries, and performing services for the caravanseries."<sup>27</sup> Thus one finds that villages also had an internal division of labor for performing non-agricultural functions in the rural area, and their obligations toward the state were not only fulfilled in kind but also in services rendered.

The average size of a village was quite large, with 500 or more houses, and a belt of such villages existed around the cities. There were also sizeable nomad groups and tribes all over the Empire, usually engaged in animal husbandry. In regions where the climate was suitable, however, their production function was twofold: animal husbandry in the summer pasture grounds and cultivation in the winter quarters. The contact of the nomad groups with the system occurred when the

<sup>23</sup> At this point, attention must be drawn to the fact that, technically speaking, only urban groups or the highly mobile merchants benefited from these services, not the villagers who mobility was much lower because they could not leave the land.

<sup>24</sup> Ö. L. Barkan, "Şehirlerin Teşekkül ve İnkişaf Tarihi Bakımından Osmanlı İmparatorluğu'nda İmaret Sitelerinin Kuruluş ve İşleyiş Tarzına ait Araştırmalar" [Researches on the establishment and the functioning of İmaret centers in the Ottoman Empire with regard to the establishment and functioning of cities], *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, vol. 23, no. 1-2 (October 1962, February 1963).

<sup>25</sup> Lütfi Gücer, *XV-XVII Asırlarda Osmanlılarda Hububat Meselesi ve Hububattan Alınan Vergiler* [The cereals question in the Ottoman Empire in the fifteenth to seventeenth centuries and taxes assessed on grains] (Istanbul, 1964).

<sup>26</sup> Zeki Arıkan, "16. Yüzyılda Isparta" [Isparta in the sixteenth century], *Belgelerle Türk Tarihi Dergisi*, no. 5 (February 1968).

<sup>27</sup> Ö. L. Barkan, "Osmanlı İmparatorluğu'nda Çiftçi Sınıflarının Hukukî Statüsü" [The legal status of the peasant class in the Ottoman Empire], *Ülkü (Halkevleri Dergisi)*, vol. 9, no. 50, pp. 102-103. Taken from S. Divitçioğlu, op. cit., pp. 14-15.

groups stopped at large cities or when temporary markets were established in summer pasture grounds.<sup>28</sup> Often large nomad groups became a part of the administrative and financial system and provided an important source of income for the state. Attempts were also made to settle the nomad groups, and the nomads who settled in cities lived in separate districts.

In this spatial organization there was a special hierarchy and integration among port cities. Because of east-west trade routes, important commercial and pirate cities and ports were seen in eastern Mediterranean regions.<sup>29</sup> Pirate ports of both commercial and military importance were well connected with their hinterlands, and they flourished during those periods when central authority weakened. In periods when the central authority gained power, such cities and princedoms submitted to the state, as they had in the sixteenth century. Both piratical and commercial functions, depending upon their radius of activity, required different sizes of ships and different levels of organizations, the spatial reflection of which was a hierarchy of port cities.

For the maintenance of such a specialized and integrated hierarchy, adequate means of communication and control are necessary. The increase of integration within the technological limits of the era could only be achieved through organization. In order to measure the degree of integration, the intensity of commodity flow and communication must be known.

The empire's important trade routes and military campaigns involving large armies required an efficient transportation organization. In 1564, the mayors (*kadis* or *qadis*) of ten cities in Bulgaria were ordered to deliver 150,240 sheep to the army at Belgrade for a military campaign.<sup>30</sup> During a campaign to the East, 40,000 *irdep* of wheat, 50,000 *irdep* or barley, and 20,000 *irdep* of beans were transported from Egypt to Tripoli and Damascus. Provisions were then transported from Tripoli to Birecik and from there to Baghdad via the Euphrates River. The provisions consisted of 99,000 camel loads, which were roughly the equivalent of 30,000 tons. Such heavy transportation took place not only in wartime but also in peace. In 1610, a caravan going from Baghdad to Basra consisted of 943 loads, or some 1,100 transport animals. Such a caravan was capable of carrying a load of about 300 tons,<sup>31</sup> and its length would be around three kilometers, which gives an idea of the size of service organizations they required at the resting points. Between Kayseri and Sivas (a distance of approximately 200 km.) alone there were 24 caravanseries.<sup>32</sup>

Security on the roads was provided by the *derbent* organization. The *derbents* were organized as villages or *timars*, and the members of the *derbent* were

<sup>28</sup> Faruk Sumer, "XVI Asırda Anadolu, Suriye ve Irak'ta Yaşayan Türk Aşiretlerine Umumî Bir Bakış" [A general glimpse at the Turkish tribes living in Anatolia, Syria, and Iraq in the sixteenth century], *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, vol. 2, nos. 1-4 (October 1949, July 1950).

<sup>29</sup> Maurice Aymard, "XVI. Yüzyılın Sonunda Akdeniz'de Korsanlık ve Venedik" [Venice and piracy at the end of the sixteenth century], *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, vol. 23, nos. 1-2 (October 1962, February 1963).

<sup>30</sup> Ahmed Refik, *Türk İdaresinde Bulgaristan 973-1255* [Bulgaria under Turkish administration 1565-1839] (İstanbul, 1933), p. 8, also see pp. 9-11.

<sup>31</sup> Halil Sahillioğlu, "Bir Tuccar Kervanı" [A trading caravan], *Belgelerle Türk Tarihi Dergisi*, no. 9 (June 1968).

<sup>32</sup> Osman Turan, *op. cit.*

responsible for the maintenance and repair of roads.<sup>33</sup> Parallel to this network of roads, there developed a communications system based on horseback messengers.

The shifting of the moveable factors of production labor also demonstrates the degree to which the system was integrated. The *surgun*, as well as being a means for achieving the best land-labor equilibrium, was also a means for ensuring the integration of newly-conquered areas into the empire by resettlement and colonization.

Population groups from integrated areas of the empire were resettled in newly-conquered areas, and population groups from the newly-conquered areas were resettled in integrated ones. In 1461, after the conquest of Trabzon, two *surgun* movements in opposite directions took place. On the one hand, people from nearby Muslim Turkish cities (Amasyan, Tokat, Samsun, Bafra) were transferred to Trabzon, and timars were distributed to *sipahis* coming from different places but mostly from Albania; on the other hand, part of the local population of Trabzon was transferred to Istanbul, and most of the former Christian *sipahis* were exiled to Rumelia.<sup>34</sup>

The shifting of the location of the mobile factor of production sectors is seen not only in agriculture but also in other production sectors. Construction works which created a temporary demand for skilled labor also depended on shifting the location of labor.<sup>35</sup> With the help of a central documentation system, it was possible to specify the number of laborers with specific skills, name by name. At extraordinary times, a great portion of the empire could send construction foremen and laborers to certain areas and thus the financial burden could also be spatially distributed.<sup>36</sup>

So far we have analyzed some of the spatial organization data indicating the degree of integration of the spatial structure of the empire in the sixteenth century. We must not forget that this integration was bounded by existing technology. The fact that a central administrative structure could penetrate such a vast area given the limited technological possibilities of the period reveals that this integration was

<sup>33</sup> Derbent villagers, like the *raaya*, could not leave the *derbent*. The only difference is that the *derbent* villagers were exempt from some taxes in return for the services they rendered. Cengiz Orhonlu, *Osmanlı İmparatorluğu'nda Derbend Teşkilâtı* [The Derbent organization in the Ottoman Empire] (Istanbul: İstanbul Üniversitesi Edebiyat Fakültesi publication, 1967), pp. 8-31, 47-55, 95-100.

<sup>34</sup> The word *surgun*, as currently used, involves the concept of punishment, whereas in the Ottoman Empire *surgun* was a planned adaptation process in spatial organization for achieving the best land-labor equilibrium as well as a means for ensuring the integration of newly-conquered areas into the empire by resettlement and colonization. Since the *raaya* could not leave the land, we cannot expect the adaptation to changes in spatial organization to be achieved through voluntary migration. Only when the Ottoman order began to decline and it was no longer possible to control the *raaya* did migration become a means of adaptation. However, at periods when voluntary migration was not possible, adaptation was achieved through forced transfer of population. See Ö. L. Barkan, "Osmanlı İmparatorluğu'nda Bir İskân ve Kolonizasyon Metodu olarak Sürgünler" [Forced migration as a method of settlement and colonization in the Ottoman Empire], *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, nos. 1-4, 11-15 (October 1953-July 1954), pp. 209-137.

<sup>35</sup> In 1583, 600 workers from Mytilene and Yörüks from Tekirdağ were brought to Istanbul in order to perform construction work. Ahmed Refik, *Türk Mimarları* [Turkish Architects] (Istanbul: Hilmi Kitapevi, 1937), pp. 111-114.

<sup>36</sup> Ö. L. Barkan, "XVI ve XVII Asırlarda Türkiye'de İnşaat İşçilerinin Hukukî Statüsü" [The legal status of construction workers in Turkey in the sixteenth and seventeenth centuries], *Sosyal Siyaset Konferansları*, fourteenth book (Istanbul, 1963).

based on a very delicate balance. Within this delicate balance local areas and regions were able to preserve their relative autonomy, although to different degrees.

The most pervasive form of control exercised by the central organization over the empire is seen in the taxation system. The documents related to taxation reveal that the relation of the capital with the states (*eyalet*) was not at all uniform. The empire had 39 states and within these states there were 176 *sancaks* and *livas*. More than 100 different taxation codes were issued for the states and sancaks of the empire.<sup>37</sup> This shows that even at the sub-state level, the type of control applied by the capital was differentiated.

With the appointment of sipahis to new timars after the conquests, ownership of the timar tended to become hereditary. When a timar owner died, the *dirlik* (literally: "means of livelihood," a general term for any grant from the Sultan to provide a living for the grantee) was usually given to his son;<sup>38</sup> moreover, after the conquest, most of the local *beys* could keep their land.

There was another informal control relation in the empire which included vast territories where total integration could not be achieved due to limited technology. The *Ahi*, a guild system, controlled production technology and set the necessary regulations. In most cases, the Ahi also had administrative and control functions. Control regulations were set by Ahi Baba, supposedly related to the Ahi Evren Dynasty, who lived in Kırşehir, and the apprentices were accepted according to his regulations.<sup>39</sup> In Istanbul another leader, Zeydi Hindi, established regulations. Informal control relations could not be established over the empire, and a special central organization was required.

As a result, we can now describe the spatial organization of the empire by asserting that it consisted of regions which, although isolated, in essence were quite integrated, given the limited technological possibilities of the time, and which preserved their autonomies and showed a distinctive hierarchy among themselves. Although there were important geographic disparities among them, they did not show marked differences of development.

#### SPATIAL ORGANIZATION IN THE SEVENTEENTH AND EIGHTEENTH CENTURIES

The loosening of the social structure of the empire during the seventeenth and eighteenth centuries was reflected in its spatial organization. The integrated spatial organization of an industrial society could not materialize, because the society did not make the transition to an industrial mode of production. Several hypotheses can be put forward to explain this failure. They may be categorized in three groups.

<sup>37</sup> Şinasi Altundağ, "Osmanlı İmparatorluğu Vergi Sistemi Hakkında Kısa bir Araştırma" [A short investigation regarding the taxation system of the Ottoman Empire], *Ankara Üniversitesi Dil ve Tarih Coğrafya Fakültesi Dergisi*, vol. 5, no. 2 (1947), p. 190.

<sup>38</sup> Cengiz Orhonlu, *Osmanlı İmparatorluğu'nda Derbend Teşkilâtı*, pp. 36-37, and Ö. L. Barkan, *Ülkü* (*Halkevleri Dergisi*), vol. 8, no. 58 (1937) and no. 59 (1937), pp. 295, 414, 420.

<sup>39</sup> Franz Taeschner, "İslam Ortaçağında Futuvva" [Futuwwa in Medieval Islam], *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, no. 15 (October 1953-July 1954).

The *first group* explains the failure to industrialize in the differences in the internal dynamics of the Ottoman social structure. These differences are basically threefold:

1. According to the arguments based on the hypothetical existence of an "Asiatic" mode of production,<sup>40</sup> the existence of a central despotic authority prevented industrialization.

2. Another reason is the "hegemonia paradox." In an empire which overexpands, effective control weakens and a substantial military apparatus becomes necessary. The structure eventually collapses under its own weight. Aspects of this hypothesis may be traced back to Ibn Khaldun.<sup>41</sup>

3. Another hypothesis is that the empire was dominated neither by a central authority nor by the feudal system,<sup>42</sup> and this situation was an obstacle to industrialization.

The *second group* seeks the reason in the internal dynamics of the European countries and their interrelations:

1. As a result of the geographic discoveries, the European countries obtained the first accumulation of capital from their new colonies. It was this accumulation of capital which helped them in their transition to the capitalistic stage.

2. Another explanation is based on the production technology of northwestern Europe. The introduction of horse-drawn plows in the tenth and eleventh centuries (at the time oxen-drawn plows were being used in the Mediterranean area) and the substitution of wheat in place of rye led to an increase of the surplus accumulating in the hands of the merchants.<sup>43</sup>

The *third group* is based on the type of relations the Ottomans had with the Western world, and thus this explanation is based on the external dynamics of Ottoman society: Capitulations and the hindrance of development by Western countries led in one sense to the colonization of Ottoman society. This reason may be valid for the period beginning with the end of the eighteenth and extending through the nineteenth century; yet, it is not an adequate explanation when applied to the seventeenth and eighteenth centuries. In this period, it could have been only a complementary factor.

The second group of hypotheses are quite acceptable as far as they go. There remains, however, the question of whether the Ottoman failure to industrialize can be attributed solely to a higher rate of development elsewhere. Did not perhaps the internal dynamics of the Ottoman society play a decisive role in delaying the development? Explanations given in the second group do not give sufficient attention to this problem. We can only assert that the "hegemonia paradox" and the failure of local groups and the central authority to maintain systematic control

<sup>40</sup> Sencer Divitçioğlu, op. cit., for a discussion on "Asiatic" mode of production, see pp. 6-26; for the "despotic central authority" (*ceberrut devlet*), see p. 63.

<sup>41</sup> For a more economics-oriented interpretation of the "hegemonia paradox" in the Ottoman Empire, see İdris Küçükömer, *Düzenin Yabancılaşması* [The alienation of the system] (Istanbul: Ant publication, 1969), pp. 41-42.

<sup>42</sup> Niyazi Berkes, *100 Soruda Türkiye İktisat Tarihi* [Turkish economic history in 100 questions], vol. 1 (Gerçek Yayınevi, 1969), p. 115.

<sup>43</sup> Mübaccel Kiray, op. cit.

prepared a suitable environment for foreign forces eventually to dominate the Ottoman Empire (by the nineteenth century) and, in the process, prevent industrialization.

Certainly, developments in Europe after the sixteenth century affected the empire by limiting its expansion, and simultaneously, the empire ceased to represent a center-front model.

In order to maintain its expansive military posture against Europe, the empire henceforth kept a large army. The permanent *janissary* army, which numbered 12,000 in the sixteenth century, rose to 200,000 excluding cadets (*ace-mioglans*),<sup>44</sup> in the eighteenth century. As a result, the military fief system lost part of its importance and state expenses increased. The change in the trade routes during the Age of Discovery caused a decrease in the income of the empire, and rising state expenses forced changes in the land tenure system. The appointment of central government officials (*kapikulu*) for the administration of timars and *zeamets* was reduced, and timars and *zeamets* (form of land grant) were sold to local lords by *mukataa*. As a result, the autonomy of local control groups increased.<sup>45</sup> During the period 1530-1580, the population of Anatolia increased 40-50%.<sup>46</sup> But the increase of population and the change in the land tenure system<sup>47</sup> disturbed the resource-population balance. Since the expansion of the empire also slowed down, this balance could not be retrieved through *surguns*. At the same time, the technology remained constant, so the balance could not be reset.

This led to the *Celali* revolts. The *raaya* revolted against the new local control system of country notables or *ayans*. Although the central authority supported the public revolts against local lords,<sup>48</sup> the latter triumphed in the end. And at the beginning of the seventeenth century, as a result of increasing exploitation, most villages were deserted, for the *raaya* had fled.

The *ayans* strengthened their control by hiring and supporting bandits. Although public uprisings and brigandage existed to a limited extent in the eighteenth century, documents show that after the second half of the eighteenth century, the brigands were controlled by *ayans*, and struggle among *ayans* was substituted for public uprising and brigandage. The *ayans* won further autonomy by a legal agreement with the central authority in 1808.

At the end of the eighteenth century, only two provincial centers were under the complete control of the capital or palace. The rest of the empire was indirectly

<sup>44</sup> İdris Küçükömer, op. cit., p. 54n., and İsmail Cem, op. cit., p. 140.

<sup>45</sup> Çagatay Uluçay, 18. ve 19. Yüzyıllarda Saruhan'da Eşkıyalık ve Halk Hareketleri [Banditism and popular movements in the eighteenth and nineteenth centuries in Saruhan] (Istanbul: Berksoy Basımevi, 1955).

<sup>46</sup> Ö. L. Barkan, "XVII. Asrın İkinci Yarısında Türkiye'nin Geçirdiği İktisadî Buhranların Sosyal Yapı Üzerindeki Tesirleri" [The influence on social structure of the economic crises undergone by Turkey in the second half of the seventeenth century], *İktisadî Kalkınmanın Sosyal Meseleleri* (Istanbul: Ekonomik ve Sosyal Etüdler Konferans Heyeti, 1964).

<sup>47</sup> The change in the control system initiated a parallel change in the agricultural crop pattern. Since animal husbandry did not require close control, most agricultural land was turned into pasture (İsmail Cem, op. cit., p. 149).

<sup>48</sup> Mustafa Akdağ, *Celâli İsyanları (1550-1603)* [The Celali revolts] (Ankara, AU DTCF publication, 1963).

administered through the channels of local authority.<sup>49</sup> Although local autonomy increased, the provinces never achieved complete independence.

The raaya, more exploited than ever, began leaving their land. These unemployed masses migrated to cities and especially to Istanbul, where the population rapidly increased. According to the historian Inciciyan,<sup>50</sup> the population of Istanbul already exceeded a million people at the end of the eighteenth century. Other sources give 873,000 as the population of 1850, and it was becoming more and more difficult to supply food for the city. Toward the end of the sixteenth century, measures were taken to slow down the migrant flow to Istanbul.<sup>51</sup> A decree was issued to turn back the incoming groups on their way to Istanbul and to send back the unemployed to their original residence.<sup>52</sup> When urbanization gains momentum without concomitant industrialization, a city cannot always provide enough employment for incoming groups. Thus, certain occupations appeared; for example, the *tablakars*, who resemble the peddlers of today. (In the eighteenth century measures were taken to prevent the increase of the *tablakars*.)<sup>53</sup> Also, the existence of the unemployed and low-income groups caused the appearance of low-quality housing resembling the shantytowns (*gecekondular*) of today.<sup>54</sup>

There was also migration to the other large cities and regional centers. Edirne, for example, which covered 350 hectares toward the end of the fifteenth century, grew to 850 hectares toward the end of the seventeenth century.<sup>55</sup> The population of Edirne was 22,335 in 1520, yet in the eighteenth century there were 20,000 households (or nearly 100,000 people).<sup>56</sup> Even if this number may be considered too high, other sources verify that the city had more than doubled.

In these growing cities the production structure was increasingly coming under the influence of Western goods. For example, in an annual fair in Bursa in the eighteenth century, inferior quality textiles from France, England, the Netherlands, and Venice imported through the Izmir port were sold and also exported to the Orient. The Western merchants were now buying primarily raw materials from Turkey;<sup>57</sup> for example, after supplying Istanbul and Aleppo textile works, about

<sup>49</sup> Doğan Avcıoğlu, op. cit., p. 33.

<sup>50</sup> P. G. Inciciyan, *XVIII Asırda İstanbul* [Istanbul in the eighteenth century] (Istanbul: İstanbul Fethi Derneği publication, 1956), p. 14.

<sup>51</sup> Rukiye Bulut, "18. Yüzyılda İstanbul" [Istanbul in the eighteenth century] *Belgelerle Türk Tarihi Dergisi*, no. 3 (December 1967), pp. 30-32.

<sup>52</sup> A good indication of the heavy immigration at that time: Evliya Çelebi mentioned that there were close to 80,000 bachelors in Istanbul (*Evliya Çelebi Seyahatnamesi*, edited and translated by Zuhuri Danişman, vol. 2, p. 26).

<sup>53</sup> İbrahim Sivrikaya, "18. Yüzyılda İportacılıkla İlk Mücadele" [The first struggles with street hawkers in Istanbul in the eighteenth century], *Belgelerle Türk Tarihi Dergisi*, no. 13 (October 1968).

<sup>54</sup> Orhan Ering, "250 Yıl Önce İstanbul'da Gecekondular Sorunu" [The shantytown problem in Istanbul 250 years ago], *Belgelerle Türk Tarihi Dergisi*, no. 12 (September 1968).

<sup>55</sup> Gündüz Özdeş, *Edirne İmar Plânına Hazırlık Etüdü* [Preparatory study for the restoration plan of Edirne] (Istanbul: İstanbul Teknik Üniversitesi Mimarlık Fakültesi Yayını, 1951), pp. 23, 28-29.

<sup>56</sup> Halil Sahillioğlu, "XVIII. Yüzyılda Edirne'nin Ticarî İmkânları" [Prospects for commerce in Edirne in the eighteenth century], *Belgelerle Türk Tarihi Dergisi*, no. 13 (October 1968).

<sup>57</sup> The accumulation of precious metals in the West increased the prices of the raw materials, so it was difficult for the Anatolian handicraftsmen to compete with Western buyers, who could afford higher prices (Niyazi Berkes, "Azgelişmişliğin Tarihsel Nedenleri" [The historical reasons of underdevelopment], *Yön* (21 October 1966).

3,000 *kantars*<sup>58</sup> of Bursa silk remained for export. Even so, at least 2,000 looms were still working in Bursa.<sup>59</sup> Good quality Ankara mohair was not then sold to Western merchants. Ankara's annual production was up to 20,000 rolls of camlet and mohair. In this period, trade with Western countries was organized through their consulates.

In the meantime, although the influence of the West was becoming stronger, regional centers were growing rapidly. They still maintained their production functions and were able to make regional controls more effective than before.

Since local autonomy now increased and central power weakened, more capital could be accumulated locally<sup>60</sup> and more surplus could remain in the locality where it was produced. This fostered the growth of regional centers, but the capital was suffering from shortages in grain supply.<sup>61</sup> Increasing taxes coupled with the pressure of the ayans also affected the smaller settlements. The villagers, abandoning their villages in valleys and on roads, took shelter in places safe from the tax collectors. Toward the end of the sixteenth century, this loosening of the central authority and abandoning of the villages resulted in important changes in the geographic appearance of Central Anatolia. In the eighteenth century, the break-up of the rural settlement pattern caused cultivation to be abandoned in many places. Those who took shelter in the mountainous areas<sup>62</sup> became shepherds, breeding enough animals to maintain themselves.<sup>63</sup>

In order to stop the brigandage<sup>64</sup> of increasing nomad groups and to resettle the abandoned areas, the empire had to make certain changes in its settlement policy. Beginning in 1691, nomad tribes were settled in abandoned areas. The aim was to reconstruct the vacant areas, to open them to cultivation once again, and to terminate the struggles between nomad tribes and settled groups.<sup>65</sup>

<sup>58</sup> 1 *kantar* is equal to 56.5 kg.

<sup>59</sup> Halil Sahillioğlu, "18. Yüzyıl Ortalarında Sanayi Bölgelerimiz ve Ticarî İmkânları" [Our industrial regions in the middle of the eighteenth century and their commercial prospects], *Belgelerle Türk Tarihi Dergisi*, no. 11 (August 1968).

<sup>60</sup> In Bursa, one rich man's wealth varied between 250,000 and 300,000 gold coins. This roughly amounted to one-ninth of the total budget of the empire. The state, which expected to collect 1,500 purses of gold from Bursa for internal loan, could only collect 100 purses (Halil Sahillioğlu, *op. cit.*). We can thus draw the conclusion that there was an accumulation of local capital due to the weakening of the central power.

<sup>61</sup> In Istanbul there was a scarcity in all the important consumption goods. For example, providing wood for the winter had become quite a problem. According to a decree issued in 1803, ports on the Black Sea coast were ordered to send 450,000 cartloads of wood to Istanbul. Many decrees were issued in the eighteenth century for the solution of such shortages. See Halil Kutluk, *Türkiye Ormanlığı ile İlgili Tarihî Vesikalar, 1487-1923* [Historical documents concerning forestry in Turkey, 1487-1923] (Istanbul: Tarım Bakanlığı Orman Genel Müdürlüğü, 1948).

<sup>62</sup> About two-thirds of the Ankara villages were completely abandoned. In Bacı county only 5 out of the former 38 villages, and in Haymana county 10 villages out of 80, remained (İsmail Cem, *op. cit.*, p. 162).

<sup>63</sup> Necdet Tunçbilek, *İç Anadolu'nun Ekonomisi Hangi İstikamette Geliştirilmelidir?* [In what direction should the economy of inner Anatolia be made to progress?], Istanbul, 1962, p. 9, taken from Doğan Avcıoğlu, *op. cit.*, p. 31.

<sup>64</sup> Çagatay Uluçay, *Saruhan'da Ezkiyalık ve Halk Hareketleri* [Banditism and popular movements in Saruhan], pp. 80-81.

<sup>65</sup> Cengiz Orhonlu, *Osmanlı İmparatorluğu'nda Aşiretlerin İskân Teşebbüsü, 1691-1696* [The attempt to settle tribes in the Ottoman Empire, 1691-1696] (Istanbul: İstanbul Üniversitesi Edebiyat Fakültesi publication, 1963), p. 5.



Although no important changes took place in the communication and transportation technologies, increasing insecurity caused a decline in central control. Insecurity and declining commercial activity reinforced each other—insecurity causing a decrease in commerce, and the very fall in the flow of commodities heightening insecurity. The changing balance in income and expenses necessitated a revaluation of expenses, and thus a decrease in the allocation of funds to charitable services, such as *imaret*. This naturally affected the facilities offered on the roads, and many caravanseries were left to decay.<sup>66</sup> Thus in this period, the integrated structure of both the interregional and regional levels loosened, the interconnection among the parts decreased, and the system became ready for a new type of integration.

In this period, we also see the appearance of nuclei within the system which in a subsequent period played an important role in the integration of the system with foreign countries. Among these we can mention the organization of foreign trade through the consulates, the organization of the religious and educational institutions of the West, and the granting of capitulations.

#### SPATIAL ORGANIZATION IN THE NINETEENTH CENTURY (Semi-Colonial Period)

In the nineteenth century, Europe was already industrialized and searching for new markets, and a consequence was that foreign influence increased in the empire. It is obvious that the empire, which was forced to become an open market after the 1740 capitulations and especially after the 1838 Anglo-Ottoman Trade Agreement, could not achieve the transition to an industrial mode of production within the matrix of such foreign political relations. With the 1838 Trade Agreement, the empire did in reality become an open market, and the production activity in Ottoman cities rapidly declined. In Bursa, the number of silk-weaving looms, had been as high as 2,000 in the eighteenth century, fell to 45 in 1845. While more than 25,000 bales of wool had been processed and exported from Ankara, in 1836 less than 5,000 bales were processed.<sup>67</sup> Moreover, after 1850, Western commercial and industrial capital began to pour in. Industries processing the first stage of raw materials were established, especially in the port cities, bearing clear witness to the increasing foreign influence rather than to the development of a national industry. The establishment of such industries necessitated the control of natural resources; the land law passed in 1858 allowed private ownership of land, and this meant abandonment of the land tenure system which had been the basis of the Ottoman order. It is now known that this law had important repercussions and paved the way for foreign influence. To facilitate the passage of this law, foreign powers extended credit to be used in cadastral applications. An 1867 law gave foreigners the right to buy land.<sup>68</sup> With the establishment of the "Council of the Public Debt"

<sup>66</sup> İsmail Cem, op. cit., p. 171.

<sup>67</sup> Alfred Bonne, *State and Economics in the Middle East: A Society in Transition* (London: Routledge and Kegan Paul, 1960), p. 231.

<sup>68</sup> In return for Ottoman debts, the British tried to colonize the Empire by buying land (c.f. Austin, *Underdeveloped Resources in Asia*, London, 1878, p. 28). The Germans pursued similar goals in the construction of the Baghdad railroad (Tevfik Çavdar, *Osmanlıların Yarı Sömürge Oluşu* [The Ottoman transformations into a semi-colony] (Istanbul: Ant publication, 1970), p. 45.

in 1881, the control of Ottoman resources by foreign powers gained in status and reached its peak.

The decline of industry, the opening of the resources of the empire to Western markets, and the change in traditional trade routes and functions led to important changes in spatial organization. The Anatolian cities thus lost the production activities and geographic specialization they had achieved in the sixteenth century and became trading centers assuming collective and distributive functions. As the empire became an open market for Europe, a new transportation system responding to the needs of the Empire's new status was born. This determined the appearance of new cities as well as the decline of others.

The most important characteristic of the cities that grew in this period was their location at the breaking points of the new transportation network. The commercial activity opportunities offered at these points encouraged their growth. Mersin, where there were a few houses in 1852, had forty years later a population of 21,756. Izmir, which was administratively tied to Aydın, became the province center in 1851. The population of Izmir rapidly increased up to 250,000 before the First World War,<sup>69</sup> and as a result, the city became the second largest urban center in Anatolia. On the other hand, the population of Bolu, which was a trade and cultural center, fell from 20,000 in 1836 to 8,000 in ten years. When the Black Sea coast was opened to trade, Trabzon, a coastal city, developed and Tokat, an internal center, declined;<sup>70</sup> later, the opening of the Suez Canal contributed to the lessening of importance of Trabzon due to changes in the Persian Trade route.

The railway network penetrated into the hinterlands of the ports to connect the two. The network's tree-like pattern is evidence of the colonial structure. This form is clearly visible in the Izmir-Aydın, Izmir-Kasaba (1868), Mudanya-Bursa (1892), and Damascus-Hama railway lines. The hinterlands where the railways penetrated were openly controlled by foreign countries. The railway system could not be internally connected due to the resistance of the foreign powers who wanted to delineate areas of influence for themselves. Growth of an internal transportation system and the delineation of areas of influence for foreign powers became identical. The Baghdad railway is a well-known example of this situation.<sup>71</sup> The Ottoman government was careful in preventing one foreign power taking over another foreign power's area of influence.

The typical settlement pattern of the era consisted of separate regions each under the influence of a certain foreign power, which controlled a port and a railway track connecting the hinterland to that port. The ports were not the only breaking points; the terminations of railway tracks served this function as well.<sup>72</sup>

<sup>69</sup> Emin Canpolat, *Izmir, Kuruluşundan Bugüne Kadar* [Izmir, from its foundation to our day] (Istanbul: İstanbul Teknik Üniversitesi, Mimarlık Fakültesi, 1953), pp. 53, 65.

<sup>70</sup> Tuncer Baykara, "19. Yüzyılda Anadolu'nun İktisaden Çöküşü ve Bugüne Etkileri" [The economic decline of Anatolia in the nineteenth century and its effects on the present], *Belgelerle Türk Tarihi Dergisi*, no. 25 (October 1969).

<sup>71</sup> M. E. Earle, *Turkey, The Great Powers and the Baghdad Railway: A Study in Imperialism* (New York: The Macmillan Co., 1924), *passim*, especially 120-146.

<sup>72</sup> This statement is inspired by William Alonso's theoretical analysis of the structure of transport costs and trans-shipment points. See his "Location Theory," *Regional Development and Planning*, eds., J. Friedmann and W. Alonso (Cambridge, Mass.: Massachusetts Institute of Technology Press, 1954) pp. 84-88.

At such points, collection systems based on a preindustrial level of technology were organized, and cities such as Afyon were able to preserve their importance. Cart drivers controlling 100 or more carts<sup>73</sup> were commonplace.<sup>74</sup>

The settlement scheme maximized the dependence of the region on external factors; and a concentration of educational and other social facilities organized by the foreign power dominated the area and helped to strengthen foreign dominance and influence.<sup>75</sup>

The importance of port cities lay in their ability to perform the collection function even without railways. A hierarchy was formed among ports since sea transportation was relatively easy. Goods were sent from small ports to large harbors on traditional small vessels and were exported from there. Because of the characteristic growth of the transportation network, this structure necessitated the formation of big harbor cities, such as Salonica, Istanbul, Iskenderun, and Beirut, in order to achieve interregional flows and relations. The organization of traditional transportation was integrated with the railways wherever tracks were paved, and the main trade routes were changed. Both caused a decline in the organization of traditional transportation and therefore created important problems. To assist the people faced with starvation in Ankara, Kirsehir, Yozgat, and Sivas (20,000 died in Keskin and 4,000 in Ankara villages), 1,700 sacks of food were sent from Izmir to Ankara. Of these, only 680 reached Ankara in 20 days and 680 sacks could be transported by 200 army mules brought to Izmir for this purpose.<sup>76</sup> These examples show a decline of the traditional transport system, as compared to those extending in the sixteenth century.

It was not only the pattern of transportation networks which prevented regional economic integration. The customs system was also developed so as to prevent this integration. After the 1838 Trade Agreement, a foreign merchant paid a 5% duty on the goods he sold in the empire, whereas an Ottoman merchant paid a 12% duty for the goods he transported from one province to another.<sup>77</sup>

Because of Istanbul's Western orientation, its acceptance of a Western type of administrative institutions, and its increasing harbor functions, the population of the city continued to increase. Although its population reached 1,200,000<sup>78</sup> in 1903, Istanbul was only the tenth largest city of Europe. In fact, the city was closer to the West than to the empire. For instance, it was easier and cheaper to import wheat to Istanbul from Europe than from Anatolia. In 1911, 21,331 ships visited the harbor with a total tonnage of 20,169,000. Annual tonnage fell to 4,500,000 in 1927.<sup>79</sup> If we consider that even in 1960 the total load carried on both national and foreign ships in all Turkish harbors was 8,200,000 tons,<sup>80</sup> we

<sup>73</sup> While camels could carry 300 kilos, water buffalo carts (*kağnı*) could carry loads up to 1,600 kilos.

<sup>74</sup> E. Banse, *Auf den Spuren der Bagdad Bahn* (Weimar: A. Duncker, 1913) pp. 132-138.

<sup>75</sup> Tefik Çavdar, op. cit., pp. 84-109.

<sup>76</sup> *Türk Ziraat Tarihine bir Bakış* [A glance at the agricultural history of Turkey] (Istanbul: Publication of the First Rural and Agricultural Development Congress, 1938).

<sup>77</sup> Doğan Avcioğlu, op. cit., p. 52.

<sup>78</sup> Curtis, William Eleroy, *The Turk and His Lost Provinces: Greece, Bulgaria, Servia, Bosnia* (Chicago, London, 1903), p. 102.

<sup>79</sup> Ahmed Hamdi, *İstanbul Limanı* [The port of Istanbul] (1929).

<sup>80</sup> *Ulaştırma (Deniz Ulaştırması)* [Communications (maritime communications)] (Ankara: State Planning Organization, 1964).

realize the empire's degree of dependence on the foreign world. Because Istanbul had such weak ties with its hinterlands, consumption goods became scarcer in the second half of the nineteenth century.<sup>81</sup> In other cities which could not import wheat from abroad due to the inadequate condition of the transport system, especially in the villages, many died of starvation in bad harvest years.

In this system of economic relations, certain social classes allied themselves with certain foreign powers. Increasing exploitation provided new economic opportunities for local minority groups. The Levantine class, more inclined to identify with Western civilization, was able to serve as a commercial intermediary, thereby gaining new power. The social pattern was also reflected in the urban scene, as a dual structure developed in the cities. On the one hand, the districts in which the elite minority and the rich lived and in which entertainment and cultural facilities were concentrated (like the Pera district in Istanbul and the Punta district in Izmir) were copied from the West. On the other hand, Turkish Muslim districts preserved their traditional urban identity and mode of life. In 1879, two separate municipalities existed in Izmir, and the port, the trolley cars, and the gas and electric services were run by foreign companies.

The relations of the empire with the West resulted in a decline of the economic life of the Central Anatolian cities. These cities became conservative centers opposing the Westernization movements as represented by the mode of life in the Western port cities.

Within this pattern of relations, the empire had to make at the same time changes in the agricultural production system in order to preserve its position as the raw material market for the West. Britain tried to organize cotton cultivation in Turkey; the Germans initiated the irrigation project for Cumra Valley in order to supply cargo for the Baghdad railway; and, around 1911, 15,073 villages in Anatolia were cultivating tobacco.<sup>82</sup>

After the contact of Ottoman society with the West, the economic structure and social stratification under the influence of foreign forces showed a development parallel to the changes in the spatial organization. Many technological developments were imported from the West, but their use was controlled by the suppliers and therefore integration in the spatial organization of the country could not be achieved. The country was divided, therefore, into several areas of influence, each integrated to a foreign power.

Port cities rapidly grew at the breaking points of this new integrated system, and Central Anatolian and Eastern Anatolian cities rapidly lost their importance. They began to decline, developed defense mechanisms against Westernization, and became conservative structures; and interregional dis-equilibrium replaced the traditional Ottoman system.

In all the regions in which capitalism had developed a market economy—like Rumelia, Marmara, Ege, and Cukurova—the feudal structure weakened, whereas Eastern and Southeastern Anatolia still kept parts of their feudal structure.

<sup>81</sup> Oya Sencer, *Türk Toplumunun Tarihsel Evrimi* [The historical evolution of Turkish society] (Istanbul: Hobora Kitabevi, 1969), p. 40.

<sup>82</sup> İsmail Hüsrev, *Türkiye Köy İktisadiyatı* [The village economy in Turkey] (Istanbul: Kadro Mecmuası Publication, 1934), p. 35.

The central administrative structure adopted from the West and the development of a communications system depending on contemporary technology could not remedy the imbalance resulting from radical changes taking place at the very foundations of the system. In 1864, there were 76 telegraph offices; in 1908, the telegraph network covered most of the empire, but even this network was not under the complete control of the central government. Before the First World War, six European countries had separate post offices in Istanbul.

We must differentiate between interregional dis-equilibrium in a disintegrated and unbalanced structure and the unbalanced but integrated structures of the West, where industrialization is achieved through market mechanisms. In the latter, there is a relationship among the sizes of urban concentrations as a result of integration. The problem is one of internal dynamics.

The interregional dis-equilibrium of the Ottoman society at the end of the nineteenth century was due to the disintegration of the internal structure and subsequent integration of each region into different foreign economies in different degrees. This is a dis-equilibrium born out of the external dynamics of the system and resulting from international imperialistic relations.

But in the nineteenth century, a reaction began against the socio-economic and spatial structure of the empire, and the people began to oppose foreign colonization. Such reaction shaped the ideology of the political movements of subsequent periods.

#### CHANGES IN SPATIAL ORGANIZATION DURING THE REPUBLICAN ERA

We can consider the spatial organization policy of the Republican era as being a reaction to that of previous eras. A basic characteristic was the reaction against the imperialistic foreign relations of the Ottoman Empire and the desire to change them, and it led to important changes in its spatial organization. In this section, we shall seek the origins of the specific policy applied in this period.

The Republican era can be studied in four periods, each characterized by different policies of economic development. Between 1923-1929, the liberal period, development was by the private sector utilizing internal resources; "etat-ism" was applied in the period 1930-1946; the period between 1946-1960 was again a period of liberal economy; and after 1960, an era of planned economy began. In all these periods no matter which policy predominated, the mixed economy always existed, and the policy of the spatial distribution of public investments always remained quite stable. The relative importance attached to the private sector determined the differences in the location of urbanization in different periods.

In the period between 1923-1946, when the intensity of reaction to the results of the Ottoman social structure remained constant, the significant national objectives from the point of view of spatial organization were: modernization, abolition of the Ottoman image, reduction of imperialistic influences to a minimum, protection of a national industry, and the development of Anatolia. These were the criteria for choices in the spatial organization of the Republican era.

The first spatial problem of the Republican era was that of resettling a rural population of 400,000 refugees coming from Greece. It is known that a population of 741,000 had been settled by 1929; this involved the resettlement of more than 5% of the total population.<sup>83</sup> After the War of Independence, 800,000 Greeks migrated from Turkey, and during the exchange agreement 150,000 more went to Greece.<sup>84</sup> The main consideration in the resettlement of this population was to make them productive as quickly as possible; so the refugees were settled in newly vacated areas. This exchange of refugees did not have an important impact on the spatial organization structure, although it did create differences in the size of the settled population and some differences in land use. Post-war demographic policy encouraged the rapid increase of population, but the government did not encourage family planning until 1960.

#### *Four Policies*

The Republican era can be summarized in terms of four main policies,<sup>85</sup> each a reaction to the Ottoman order:

1. The transfer of the administrative and cultural center from Istanbul to Ankara.
2. The transformation of many provincial centers into modern administrative and cultural centers, thus introducing and initiating social changes to their hinterlands.
3. The construction of a railroad network to cover the country and thus replace the former "tree form," and the development of highways to support the railways.
4. The location of large public industries in small Anatolian towns.

The transfer of the capital from Istanbul—which had been the capital of Byzantium for more than 1,000 years and of the Ottoman Empire for more than 500 years—to Ankara<sup>86</sup> was not only born from a desire to continue the image of the War of Independence (between 1920-1923) and to locate the government in the heart of Turkey, in Anatolia; but it was also a revolt against the way of life of the Levantine class of Istanbul who identified themselves with the West. Behind the competition of Ankara with Istanbul there was the desire to create a new bourgeoisie of a Republican character that would counter the Levantine groups of Istanbul.

Placement of the country's decision-making functions in Ankara was the most important factor in causing self-perpetuating growth; concentration of higher educational functions also furthered the development of Ankara. Within the new

<sup>83</sup> *İşkân Tarihçesi* [History of settlement] (Istanbul, 1932).

<sup>84</sup> Cevat Geray, *Türkiye'den ve Türkiye'ye Göçler ve Göçmenlerin İşkânı* [Turkish out- and in-migration and the settlement of migrants] (Ankara: Siyasal Bilgiler Fakültesi, 1962), pp. 3-4, 48.

<sup>85</sup> M. D. Rivkin, *Area Development for National Growth: The Turkish Precedent* (New York: Praeger Special Studies, 1965), pp. 48-74.

<sup>86</sup> This was the first example in the world of the transfer of a capital from the largest city to a smaller inland city. The example has been repeated in several countries. Moreover, it is a very early example of the successful creation of growth poles through a regional settlement policy.

system, the tensions created between Istanbul and Ankara were soon dissolved and Ankara was integrated into the system.

Provincial centers were to be developed as cultural centers from which reforms could penetrate the hinterland. This was how the Western mode of life, perceived differently in the Republican period, was to be introduced. The Republican bourgeoisie was going to be born. The "people's houses" (*Halk evleri*, local cultural centers) are the result of such a program and thought. Besides modernizing society, the people's houses were to provide for the cultural integration of a disintegrated society inherited from Ottoman times.<sup>87</sup> Thus, Anatolian provincial centers were used as contact points in introducing social change, and they became a means to destroy the conservative centers established in the former period.

The slow rate of urbanization also made the municipal efforts toward the "beautification" of cities possible. The reconstruction of Ankara became an example for such activities in other cities. An important part of the green areas and parks in the cities were built in this period.

The major policy of this era in the field of transportation was that of railways. The main objectives after the War of Independence were the repair of tracks destroyed in the war, nationalization of tracks held by foreigners, and construction of new railroads. The main reason for insistence on this policy resulted from the bottlenecks met with during the First World War and then during the War of Independence.<sup>88</sup> Railway policy was not implemented according to a definite program. Main lines in Central Anatolia in the East-West direction formed a network, and this network was tied to the port cities. A highway system was then developed to support the railways.

The fourth settlement policy, the location of industry in the interior, drew the most criticism. The policy could now be implemented because of the improvement in the railway network, thus illustrating the internal complementarities of the four policies. The major issue was not the choice of location but whether development would be realized by the public sector or by the private sector.

The factories established had other aims than just production. One aim was that a factory should be a focal point of social change in the area; another was that it should encourage the development of the private sector and thus initiate self-sustaining growth. Before discussing whether or not the aims were realized, we must consider whether it is realistic to expect the fulfillment of these goals.

Since the profits of these factories were sent to their central offices and from there distributed to new investment areas, the factories could only have indirectly developmental effects on the areas through the consumption activity of the workers. So we must direct our analysis to both forward and backward linkage effects. Because most of such factories produced final consumption goods (for instance, textiles and sugar), they could stimulate the production of raw materials such as sugar cane, but the factories could not be expected to attract other

<sup>87</sup> R. D. Robinson, *The First Turkish Republic* (Cambridge, Mass.: Harvard University Press, 1963), pp. 37-92.

<sup>88</sup> Murat Ergun, *Bir Demiryolcunun Kurtuluş Savaşı Hatıraları* [Memoirs of a railwayman concerning the War of Independence] (Istanbul, 1966).

industries. There are no substantial sociological studies showing the effect such factories have had in introducing social change to their surroundings, and although their isolation from their environment and the restriction of their social services to small groups would serve as limiting factors, there is reason to believe that they were not without an impact.

In addition to these four settlement policies, another important policy influencing spatial organization was that of encouraging a population increase. This policy was considered rational in the first years of the era because of the low density and the possibilities for opening new areas to agriculture. Health programs were implemented to support the policy, such as that of malaria eradication—one of the most successful campaigns in the world—which turned the study of water and land resources into a marsh-drying project (e.g., the drying of Cellat Marsh) rather than irrigation and energy production. The study of resource development remained limited and could not be realized despite important investment allocation.<sup>89</sup> The rate of urbanization was quite low because agricultural land was still available, mechanization was not widespread, and industry in small cities had not yet reached a critical size. This was more in line with the policy of holding rural population in the agricultural areas.

At the end of this era we can see the change in many structural factors of the previous era. Special care was given to Central Anatolia in an attempt to remedy interregional disparities, but Eastern Anatolia remained neglected. The rates of growth of Istanbul and Izmir were below the rate of increase of the total population and the total urban population. On the other hand, Central Anatolian cities, led by Ankara and Eskisehir, grew rapidly<sup>90</sup>—the result of a spatial organization during this era which was oriented toward achieving an interregional balance.

After the Second World War, there was a change in the economic development policy. The policy of development by the public sector with the use of internal resources was abandoned, and development was to take place by using foreign aid and through the private sector. Although there was not a great change in the shares of the public and private sector of the economy, the system of relations was changed. A policy of development through agriculture, rather than industry, and increased emphasis on infrastructural investments was chosen. This new policy had important effects on spatial organization, such as the expansion of agricultural land through mechanization and the shift of priority from railway to highway investments.

Beginning in May, 1949, tractors arrived through the Marshall Aid Program. In accordance with the recommendation of the U. S. Military Aid Organization in 1947, a large highway investment program was undertaken by the General Directorate of Highways (the Directorate had been established with the guidance of the U. S. Public Road Administration). In 1948, the Iskenderun-Erzurum road was

<sup>89</sup> Of the 100 million Turkish pounds allocated in 1929 to be spent in 12 years, only 12 million Turkish pounds were spent eight years later. Ahmet Demir, *Türkiye İç Sularından Faydalanma* [Taking advantage of internal waters of Turkey] (Ankara: Siyasal Bilgiler Fakültesi, 1963).

<sup>90</sup> Ruşen Keleş, *Şehir ve Bölge Planlaması Bakımından Şehirleşme Hareketleri* [Urban development from the point of view of city and regional planning] (Ankara: Siyasal Bilgiler Fakültesi, 1961).





Fig. 10. Turkish cities A.D. 1935.



Fig. 11. Turkish cities A.D. 1965.



Fig. 12. Turkish cities A.D. 2000.

The rapid increase in size and number of communities qualifying as "urban" in Turkey is illustrated by these three maps, two historical and one conjectural. The map of urban communities of Turkey in the year 2000 was prepared by population projection techniques. It was assumed that the national population will grow from its present 36 million (1970) to 76 million by the year 2000; that the proportion of the population living in rural communities will continue to decline approximately as in the past; and that recent rates of growth of particular communities, as well as historical patterns of city growth in Turkey, in general, are indicative of which communities will grow more than others. Unanticipated economic, administrative, and demographic developments will undoubtedly intervene and require modification of the projection in many details. Nevertheless, the general pattern of urban Turkey in the not distant future may well emerge approximately as sketched here. (Courtesy: Frederic C. Shorter.)

built for military purposes. This road program did not take into consideration the integration of highways and railroads. In the first years it had an encouraging effect on the railway freight but later the unplanned competition caused an economic crisis among the railways. Agricultural mechanization<sup>91</sup> achieved with the assistance of foreign aid not only increased cultivable land but also caused the unemployment of an important portion of the labor force. Population growth and agricultural unemployment combined with an improved highways system resulted in a rapid urbanization. Rapidly growing rural-urban contacts strengthened the market economy in the south and west coastal regions, and increased urban-rural integration.

<sup>91</sup> *Türkiye'de Zirai Makinalaşma* [Agricultural mechanization in Turkey] (Ankara: Siyasal Bilgiler Fakültesi, 1954).

In a period when more importance was attached to infrastructural investments rather than to productive investment, priority was given to the development of water resources and highways. The water resources development program emphasized the building of dams<sup>92</sup> rather than irrigation. The increased allocation of infrastructural investment meant a decrease in the share of industrial investment coming from the public sector. Although industrial investments in the public sector decreased generally, their allocation to Central Anatolia continued as a political maneuver to secure votes. The importance of port cities increased once again (like Istanbul and Izmir), and Western and Central Anatolian cities with a population of 50,000 continued to grow, whereas urbanization remained quite low in Southern and Eastern Anatolia.

After the Revolution of 1960, which occurred at the end of an inflationist era, there was a return to the policy practiced during the first years of the Republic. Again the question of interregional balance gained importance, and for the first time developmental priority was given not to Central Anatolia but to Southeastern Anatolia. With the return to power of the Democratic Party, as the new Justice Party, the policy of the First Five-Year Plan was soon changed and a Second Five-Year Plan was developed from it. The policy of using urbanization as a means of development gained importance along with the policy of interregional balance. After the Second World War, and despite the impact of the 1960s, the social and economic structure continued expressing the same characteristic and decisive influence on the formation of the spatial organization. The only change brought about during this period was the encouragement of urbanization in the east.

### CONCLUSION

It is important to consider the question of changes in the spatial organization in historical perspective. Such an approach aids in clarifying the more significant variables of the system. When the Turkish historical experience is studied from the viewpoint of changes in spatial organization, the decisive variables are the level of technology, the basic institutions of social structure, and the external relations of the system. There is a one-to-one correspondence between the changes in the external relations and institutions and the changes in the spatial organization. Of several variables, only technology remained constant in the sixteenth, seventeenth, eighteenth, and even in the nineteenth centuries. Changes in the social institutions in the seventeenth and eighteenth centuries and changes in the matrix of foreign relations also had their divergent impacts on the spatial organization. Only in the twentieth century did changes in technology, foreign relations, and social institutions converge as mutually-interacting variables.

What is the theoretical relevance of the above discussion? Where do the present location theories fit into such a framework? Location theories, it might be argued, consider the variables mentioned in the foregoing paragraph as constants. And, as a

<sup>92</sup> Between 1953-1962 about 60% of the investments of the state water works department went into dam building (Ahmet Demir, *op. cit.*).

result, the solution to problems using these theories cannot go beyond increasing efficiency in a given system. When the question is that of planning structural changes in the society, present location theories and analysis techniques will not suffice. It is necessary, therefore, to develop a macro-location theory establishing the links between basic social institutions and spatial organization. The independent variables of such a theory will be population, technology, and institutional arrangements.

## GENERAL BIBLIOGRAPHY

- Avcıoğlu, Doğan. *Türkiye'nin Düzeni* [The system of Turkey]. Ankara: Bilgi Yayınevi, 1968.
- Bonne, Alfred. *State and Economics in the Middle East: A Society in Transition*. London: Routledge and Kegan Paul, 1960.
- Çavdar, Tevfik. *Osmanlıların Yarı-Sömürge Oluşu* [The Ottoman transformation into a semi-colony]. Istanbul: Ant Yayınları, 1970.
- Earle, M. E. *Turkey, the Great Powers, and the Baghdad Railway*. New York: Macmillan, 1924.
- Eldem, Vedat. *Osmanlı İmparatorluğu'nun İktisadi Şartları Hakkında bir Tetkik* [A study on the economic conditions of the Ottoman Empire]. Ankara: Türkiye İş Bankası Kültür Yayınları, 1970.
- Hershlag, Z. Y. *Turkey: The Challenge of Growth* (2nd revised edition). Leiden: Brill, 1968.
- Karpat, K. H. *Turkey's Politics: The Transition to a Multi-Party System*. Princeton: Princeton University Press, 1959.
- Lewis, Barnard. *The Emergence of Modern Turkey*. London: Oxford University Press, 1961.
- Mears, Grinnel Eliot. *Modern Turkey*. New York: Macmillan, 1924.
- Mustafa, Akdağ. *Türkiye'nin İktisadî ve İctimaî Tarihi* [Economic and social history of Turkey]. Ankara: Türk Tarih Kurumu, 1959.
- Rivkin, M. D. *Area Development for National Growth*. New York: Praeger, 1965.
- Robinson, R. D. *The First Turkish Republic*. Cambridge, Mass.: Harvard University Press, 1963.
- Taner, Timur. *Türk Devrimi ve Sonrası, 1919-1946* [The Turkish Revolution and its aftermath, 1919-1946]. Istanbul: Doğan Yayınları, 1971.
- Uzunçarşılı, İ. H. *Osmanlı Tarihi* [Ottoman history], vol. 3. Ankara: Türk Tarih Kurumu, 1954.

---

# 31 Principles of Planning and Zionist Ideology: The Israeli Development Town

Elisabeth A. Altman and Betsey R. Rosenbaum

## Introduction

In 1948, the United Nations proclaimed the establishment of the State of Israel. Israel immediately became engaged in a war that demonstrated the difficulties the newly founded country would confront in maintaining its sovereignty. The goal of building a state and a Jewish homeland was supported by several positive resources, but these contributions proved to be problematical. The assets included a newly legitimized government, unsettled land, and a heavy infusion of capital (largely foreign) and labor. For policy-makers and planners, however, there was the need to utilize these resources at once and confusion over how to use them. To be precise, there were three major internal problems: (1) dispersing the population; (2) absorbing the immigrants; and (3) providing for security and defense.

In the abstract, planning principles should determine the selection of implementation policies. In this case, the well-articulated Zionist ideology had already established the ideal social image toward which Israeli society was to strive. Ideology is an independent variable, and the functions of planning and the national strategies employed may depend on ideology for their direction. It is possible to imagine how ideology might bias planning by predisposing individual planners toward one course of action rather than another. In this article, we must be mindful of this phenomenon and seek to understand the extent of the influence of ideology upon planning.

The focus of this article will be the development towns of the hinterlands—how these urban settlements were planned and how they evolved within the predominant ideology. The conflict and coexistence of ideology and planning principles can be seen in the use of development towns as a primary mechanism for pursuing the three objectives mentioned above. Furthermore, an analysis of national policies and conditions surrounding the policymaking process illustrates how an agrarian ideology first molds and then adapts to a plan for urban settlements. In order to carry out this discussion, this article will examine the functions and tenets of Zionist ideology, the composition of Israel's population, and the diverse expectations placed upon the concept of a development town.

What do we mean by ideology? Ideology is a definition of reality which is rooted in a historic past and which draws a clear picture for men in the

present of what society should become. An ideology has a set of political and/or social principles (ideals) which not only order thought, but also direct political or social action. Ideology is characterized by its substantive content, its ethical role in a contemporary social setting, and its method of operation.

In the latter half of the nineteenth century, the rise of Zionism, stimulated by other European nationalistic movements, represented the affirmation of the cultural and national identities of the Jewish people. Western European Jews were more affected by the Enlightenment and assimilationism within their particular countries of residence. The Eastern European Jews, however, were constantly reminded of their "separateness" by the restrictions of the governments of their "temporary homes." Zionism emerged, therefore, not because of conscious anti-Semitism (which came into play in later historical developments) but as a basic fact of life for a cohesive ethnic group living "separate lives" in Eastern Europe. One might even say that it was a reaction to the dangers of assimilation which the Eastern European Jews recognized in their Western European brothers. "Zionism is the specific form Jewish nationalism has assumed in relation to the land of Israel. Essentially, it stands for affirmation of Jewish group identity, on conditions of security and dignity in a country in which they form the majority . . ." (Akzin and Dror, 1966: 34). One should be explicit about both what an ideology includes and what it excludes. For example, Zionist ideology is a Jewish phenomenon which ignores the existence of an indigenous Arab population.

Attachment to the very earth of Palestine is one of the basic tenets of Zionism. The movement "back to the land" refers both to the spiritual attachment to historical landmarks and also to a practical desire to reacquaint Jews with working the land as farmers. The agrarian bias is a result of attempts to overturn the occupational pyramid which Jews had found for themselves in Europe, i.e., unable to own land in Eastern Europe, Jews became merchants, academicians, professionals (when they were able to enter universities), and small tradesmen. This was an attempt to "normalize Jewish life" by expanding occupational interests and opportunities. This resulted in preferential advantages and status given to those newly arrived Zionists who devoted themselves to manual labor and to an agrarian lifestyle. It also resulted in a marked, self-conscious deemphasis and denigration of the urban life which they had known (Cohen, 1970b).

Another tenet of Zionism was the practical effort to provide an economically independent society in Palestine, and later in their own sovereign state. Conquest and cultivation of the neglected desert areas were major objectives of the movement.

The socialist branch of Zionism developed from the Jewish socialist societies of Russia and other parts of Eastern Europe in the late nineteenth and early twentieth centuries. The second major wave of immigration (called the Second Aliya, or rising up) was made up mainly of such socialists who set the tone for the "pioneer image" in Israel. Self-sacrifice (even to personal depri-

vation), delayed gratification for the good of the future society, the principle of collective decisionmaking and production, and a focus on an idealized future society characterized the pioneering Zionists (Eisenstadt, 1970a:673).

The success of these pioneers in establishing a viable social structure was attributed to the force of their personalities, but even more so to their ideology. Adherence to Zionist principles could, in effect, work miracles, molding people and society to a certain form. Since human beings give ideologies their vitality and since the majority of political, cultural, and social leaders were ardent Zionists, it is not difficult to understand how Zionism became the public ethic. The rationale for this public creed has been that if pioneering Zionism could demand self-sacrifice from its original adherents, it could also make identical demands on subsequent immigrants, despite the fact that most Israeli immigrants have come for highly personal reasons often unrelated to Zionist principles.

In time, however, ideology has become routinized: Interest groups pay ritualistic lip service to it while political opinions and debates are still couched in strict ideological terms. However, in the actual functioning of Israeli ideology today,

the significance of Zionism as the guide to personal action has diminished though by no means disappeared. It still operates . . . mainly on the institutional level, providing accepted goals for public and political activity, to be achieved more through bureaucratic methods than through individual involvement by the population at large (Akzin and Dror, 1966).

The ideals of pioneering, utopianism, and egalitarianism now compete with the rising middle-class ethic and cultural diversity, relatively unknown before the founding of the state. Consequently, some of the strategies employed to meet ideological goals may not be acceptable to the original ideological principles. The use of development towns, for example, illustrates how ideology offers goals but not entirely adequate means of implementation.

### **Israel's Immigrant Population**

Prior to the establishment of Statehood, immigration of Jews into Palestine had been restricted. "In its very Declaration of Independence, Israel repealed all limitations upon Jewish immigration and later formalized, in the 1950 *Law of Return*, the principle right of all Jews to immigrate to the country" (Matras, 1965: 33). Not only was the country confronted then with the task of providing the basic necessities of food, clothing, and shelter to thousands of immigrants, but it also needed to devise a policy for dispersing the immigrants into the interior. Of necessity, the policy had to allow for the reclamation of the land for economic and defense purposes. The success of a policy directed toward these ends would also depend on the country's ability to absorb the immigrants into the general social configuration of the society.

It is important to examine what the characteristics of immigration were before and after 1948. Prior to 1948, the dominant waves of immigrants were of European origin (Cohen, 1970a: 12). They had come mainly from Russia, Poland, Austria, and Germany. In contrast, the post-1948 influx was substantially different. While a significant number continued to arrive from Europe, an increase of Afro-Asians and Orientals outnumbered the dominant European population. Table 1 illustrates this point.

Table 1. Immigrants by Country of Origin and Period of Immigration

| Period of Immigration | Absolute Numbers |                 |           | Percentages  |                 |       |
|-----------------------|------------------|-----------------|-----------|--------------|-----------------|-------|
|                       | Africa, Asia     | Europe, America | Total     | Africa, Asia | Europe, America | Total |
| 1919-1948             | 44,809           | 385,066         | 452,158   | 10.4         | 89.6            | 100.0 |
| 1948-1962             | 575,755          | 479,605         | 1,074,792 | 54.6         | 45.4            | 100.0 |

Source: Fein, 1967: 37.

Aside from the demographic characteristics of these two immigration waves, other factors also demonstrate differences. For example, the majority of immigrants prior to 1948 were young adults, either single or married, who had come to work on the land, as agricultural settlers. In contrast to this, the post-1948 immigrants included, among the Europeans, a considerable number of people, ravaged by the war, who needed care. The Afro-Asians and Orientals, on the other hand, could be characterized by the predominance of large-sized families, strong clan ties, a different cultural experience, a lower cultural and economic level, and a different religious orientation. Table 2 illustrates these characteristics.

Table 2. Family Size of Immigrants (Percentage)

|                        | Pre-1948 | Post-1948 |
|------------------------|----------|-----------|
| Single people          | 30.7     | 17.3      |
| Children               | 20.8     | 11.3      |
| Families with 5 people | 19.3     | 37.3      |

Source: Spiegel, 1966:12

Still another contrast between immigrants prior to 1948 and those who arrived after 1948 is the factor of motivation. Judith Shuval, in her study of immigration patterns into Israel, notes: "On an overall level, the basic motivating force to immigrate on the part of post-1948 immigrants was probably not so much an internal ideological one, which was characteristic of pre-1948, but an external situational one" (Shuval, 1963: 46). The significance of the Vatikim (old pioneer or veteran) in Israeli society was closely tied with Zionist ideology. As pioneers, the Vatikim had made decisive contributions to the pre-State, as farmers and as fighters. They were imbued with Zionist fervor and deeply attached to the ideology of Zionism. In juxtaposition to the Vatikim were the Olim, the greenhorns, who neither shared the attachment to the land, nor the fervor and spirit of Zionist pio-



neers. The Olim had sought out Israel because it was a homeland for the Jews and a haven from anti-Semitic sentiments (Spiegel, 1966: 28).

The precipitous migration of hundreds of thousands of immigrants required that the country devise a policy to blend the contrasts inherent in pre-1948 and post-1948 immigration. Israel, if it were to survive, had to establish a form of social integration which would create a common experience and form a basis for the organic continuity between the past, the present, and the future. The government also needed to develop, populate, and provide employment for the entire territory in order to achieve a viable nation. The salient question is: Could ideology make a significant contribution to these objectives?

### **The National Policy of Population Dispersion**

The vehicle through which the Israeli government planned to implement its objectives of (1) security and defense, (2) absorption and integration of immigrants, and (3) economic development was incorporated in the national policy of population dispersion. The central rationale for this particular approach concurred with Zionist ideology, as mentioned earlier. Coupled with this ideological preference was the very real value accorded agricultural labor. "This toil is only to be understood if taken as the expression of an eschatological faith in the healing power of peasant labor as the only way of regaining the closest and most tender union with the holy soil of the Bible" (Spiegel, 1966: 85).

The agricultural orientation which the policy of population dispersion embraced was also a reflection of previous experience in planning. Prior to 1948, there were few plans or policies for urbanization. Instead, most national efforts had focused on agricultural development in one form or another (Schacher, 1971: 363). Thus, when Israel recognized the prospect of responding to the requirements provoked by large swells of immigration, the only planning background it had was based on past experience with agricultural settlement. The consequences of this particular experience meant that future planning efforts would be influenced by this bias. Furthermore, effort would have to be directed toward preventing further settlement in the urban centers of the coastal plain, in order for occupation and development of the hinterland to succeed.

Despite national efforts at rural development prior to 1948, most of the immigrants had settled in this central zone. In fact, as of November 11, 1948, 77.5 percent of the Jewish population was settled on 11.1 percent of the land, encompassed in the Tel Aviv, Haifa, and Jerusalem metropolitan area (Spiegel, 1966, p. 2). The policy of population dispersion, therefore, meant that the two predominant tendencies of the veteran population (preference for the fertile coastal areas and the large cities) would have to be reversed for the new immigrants (Spiegel, 1966: 3).

Acclimatizing the new immigrants to a rural orientation was as considerable a task as reversing the veteran population's predisposition to urban

living. The new European immigrants traditionally were already integrated into the process of urbanization. They had concentrated their settlements, accordingly, in the relatively large European urban communities. The settlement patterns of Jews in Moslem and Oriental countries were also predominantly urban, especially among the wealthy Jewish elite. The impoverished Jews' style of life was more similar to that of the indigenous non-Jewish population (Berler, 1970).

In view of the apparent dichotomy between the immigrant tradition of urban living and the government's policy of population dispersion into the rural areas, it is necessary to examine the manner in which this policy was implemented. Of primary importance to facilitating this policy was a systematic method to guide controlled migration into sparsely settled areas. Under the auspices of the Department for Colonization of the Jewish Agency a policy of direct transfer of immigrants from the "boat directly to the hinterlands was instituted" (Cohen, 1970a: 596).

It was felt by the Department of Colonization that unless direct control was exerted, there would be a possibility that the economically or culturally more advanced elements among the new immigrants might "escape" into the established urban centers. If this occurred, the loss of human resources and vitality might slow the development of the hinterland (Cohen, 1970a: 596). For those who lacked such economic resources, it was felt that they would not object to such control because most immigrants supposedly had no preferences for location, other than to live in Israel. In actuality, this assumption was not completely true. Many newcomers wished to live near family or friends who had come before them—at least with others from the "old country."

The selection of a national policy of population dispersion concurred with Zionist ideology. The incongruity which existed between the reality of immigrant experience in urban living was outweighed by the planners' commitment to developing the land. The very essence of the policy's acceptance as a planning objective was linked to the ideological maxim of "beginning a new way of life in Israel" (Cohen, 1970a: 596). Furthermore, the agency, while lacking a clear-cut image of what kinds of criteria for immigrant selection would best facilitate the utilization of human resources within the country, apparently felt that even if they had a realistic image, it would probably be useless because they could not predict the immigrant waves.

#### **Planning Theory in Israel**

It was necessary to devise a series of plans which would incorporate the realities of Israel's physical structure, natural resources, and development demands. For example, because of limited natural resources, an agro-economy would be unable to support a mass influx of immigrants. It was also evident that amassing new populations in the existing cities would create an untenable situation (Berler, 1970). The already crowded conditions in urban areas strained the quality and quantity of human services. Furthermore,

there was a greater predisposition to ignore rather than to solve the urban problems in the older, burgeoning cities. It was obvious, therefore, that there was a need for urbanization and industrialization in the outlying regions of the country.

The national physical plan searched for a way to conform with the principles of the policy of population dispersion, the realities of the Israeli situation, and the original tenets of Zionist ideology. The plan called for the creation of a new concept of settlements in sparsely inhabited areas of the country. These settlements were designed to achieve a more balanced geographical distribution of the population. The term which was used to identify this concept was "development towns." It connoted special physical planning areas which were to be established from development budget funds in target zones of special concern to the State. As the program for development towns was operationalized, its definition was broadened to include the majority of towns founded after the establishment of Israel (Cohen, 1970a: 587).

The decision to accept the development town concept as the principal *modus operandi* for dispersing the population was not made without considerable debate among various policymakers, citizens, and planners in the new country. According to Eliezer Brutzkus (1964: 17), as early as 1936, a group of regional planners had recognized the need to achieve far-reaching structural changes in the settlement patterns within the country. Such concerns as this emerged in the form of two schools of thought, each with a model to guide the physical parameters of colonization of the hinterlands.

One school of thought, influenced by Kibbutz and Moshav leadership, held that small rural settlements (*moshavs*) should continue to be developed in a polar pattern which would link them directly with the major metropolitan centers. They felt that mass migration should continue to take root in the development of an agro-economy based in small rural settlements.

The proponents of the polar model based their conclusions on the following reasoning: Since a highly developed and centralized pattern of communications and market distributions was contained in the large urban centers, and whereas technological advances would continue to facilitate direct linkages between the rural settlements and urban centers, there would, therefore, be no justification for developing future urban centers in the hinterlands.

The second school of thought considered the polar pattern suitable for the primary and pioneering efforts of development only and held that the polar pattern was undesirable for the country's more advanced development. This school of thought "having familiarity with the different prototypes representing the spatial distribution of urban populations throughout the world, looked outside the physical realities of Israel to other models" (Brutzkus, 1964: 4).

The influence of British and European planning concepts had already been incorporated into early planning efforts initiated during the British

Mandate in Israel. This theoretical perspective led them to conclude that the spatial distribution of the country was incomplete. While the country had developed both small and large communities, it had neglected the development of intermediate-sized towns. To alleviate this situation, the planners turned to Christaller's Regional Hierarchical Model as a more serviceable prototype for physical development in the interior.

The model rationalized the opportunity for development in a manner different from the already existing polar conditions with their rural-urban linkages. Brutzkus noted (1964: 19) that the Hierarchical Model establishes urban and semiurban centers as intermediary links between the small settlements and the large cities.

To a certain extent the model abstractly reflected the traditional settlement patterns of European countries. However, the aim of this model was to replace, at least partially, the direct link between the rural communities and the three largest Israeli cities. The model would provide a more precisely defined settlement pattern, where economic and social connections within smaller regional units could emerge.

The proposed Hierarchical Model for Israeli settlement patterns appeared in the form shown in table 3.

Table 3. Urban Settlement Hierarchy

| Type of Settlement            | Population       | Characteristics   |
|-------------------------------|------------------|---|
| A. Village, kibbutz or moshav | Up to 500        |   |
| B. Rural center               | 500-1,000        | Economic, social, and cultural center for 4-6 moshavs. Residents to be craftsmen, service providers.  |
| C. Rural-urban center         | 6,000-12,000     | Serves an area with a radius of approx. 10 miles, including 30 villages. To include secondary schools, housing for farm laborers, rural-related industry, some administrative services. |
| D. Medium-sized town          | 15,000-60,000    | Regional center for commerce, culture, administration, industry locationally independent.   |
| E. Large city                 | 100,000 and over | Major central functions of government, education, and commerce.   |

Source: Strong, 1971: 172.

In the opinion of the planners, the Urban Settlement Hierarchy "appeared to be more mature and adequate for a country which was to become a densely populated and intensely developed one" (Brutzkus, 1964: 17). The plan also had other advantages. Immigrants had to be settled, and money was scarce. Land in the urban areas was privately owned and relatively expensive, while outlying land was in the public domain and, therefore, free for government

development. For defense purposes, people had to be settled near the borders, and because a rural density was insufficient, towns had to be built (Strong, 1971: 172). Furthermore, such colonization of the land provided a method to preserve close links with it.

Thus, from these prototypes of spatial development and the circumstances of Israeli reality, the planners drew these observations into a somewhat disjointed plan. The plan's acceptance did not occur without opposition. Those who favored the polar model failed to see the wisdom of the plan. Others claimed that the hierarchical model ignored the potential of technology as a key variable in eliminating the creation of new urban centers in the hinterlands. Since the model did account for new urban settlements, there was opposition from some rural spokesmen who were ideologically disapproving of city life (Cohen, 1970b). Conversely, others felt that the model ignored the antirural sentiments and traditions which were exhibited by the majority of the Jewish population.

Despite generally adverse public opinion toward the hierarchical model, the plan was quietly accepted through informal channels. The Planning Department submitted the plan to the Labour Ministry, directly responsible for the housing of immigrants. Here the plan underwent some minor revisions, but was soon adopted. Eventually it was accepted by the Prime Minister as the vehicle by which the policy of population dispersion was to be achieved (Brutzkus, 1964: 18).

### **The Experience of the Development Towns**

How well did the hierarchical model meet expectations? It is evident that successes in implementing the hierarchical model would offer better prospects for the physical and social establishment of the individual new towns and vice versa. This intricate interrelationship cannot be stressed too heavily. The following discussion of the development towns will demonstrate the need for, and finally the reluctant ideological acceptance of, more organic, regional planning methods. This examination should encompass the physical plans, the tasks which the development towns were to accomplish (e.g., absorption of immigrants, dispersal of the population), the social trends and demographic characteristics of the population, and the contributions of the development towns to the Israeli economy.

The seven Master Plans spanning twenty years gave rise to some thirty new towns demonstrating progressive phases in planning thought and implementation. The agencies (public and private) responsible for constructing the new towns include the Ministry of Housing, the Jewish Agency (responsible for the location of immigrants), and various other ministries concerned with the mix of public and private investment and administration of the towns. Centralized national planning was a particularly weak phenomenon in Israel. There was little coordination in goal setting, planning, and implementation, but there was a consensus of opinion that the development towns would act

as service centers for the villages (mainly kibbutzim and moshavim—the collective agricultural settlements) in their respective regions.

The early development towns were placed in the midst of these collectives to provide consumer, trade, and cultural services. However, adequate links between the new towns and the already existing and self-sufficient infrastructure built up among the collective villages were not established (Cohen, 1970b). There was indifference or antipathy from the rural inhabitants toward the new “planted” urban centers. For the distribution of their own agricultural products and for certain health and consumer services, the collectives communicated directly with their head offices in Tel Aviv or Haifa. Thus, the development towns (both small and middle sized) were bypassed by the villages since the early development towns could offer them only few social and economic reasons for establishing interdependent relationships.

One significant economic link, however, was that the towns became sources of hired labor for the collectives. During the early phases of the foundation of development towns, much of the industrialization outside the major urban areas (Tel Aviv, Haifa, Jerusalem) was comprised of light industries located in the collective villages or in other vaguely defined regional industrial centers—distinct from the original towns. This occurred despite initial ideological opposition to such industry requiring the hiring of outside laborers who would not share in the benefits of production (Spiegel, 1966:81–82).

The confusion of “development town” and “regional industrial center” is more than just theoretical, and it results from a self-conscious attempt to preserve the rural lifestyle of the hinterland. “It looks rather as if the rural hinterland by its own efforts as well as with the help of influential quarters, is striving for autonomy in this field too, leaving the towns entirely dependent upon themselves for their economic development” (Spiegel, 1966:82). Practically speaking, this situation is self-defeating because the development towns were supposed to function as service centers, and, consequently, they had no economic base of their own. The design of the new towns was to be modeled after the garden cities of the European and English varieties (Brutzkus, 1964: 64). The appeals of the garden city concept, with its emphasis on semirural habitations, was not dissimilar from the ideological proclivity to return to the land. Low densities, open space and greenbelts, and auxiliary farming areas were part of the model.

There are several reasons why this plan was unsuited to the Israeli experience. First, the arid climate was not conducive to the lush green areas which the planners envisioned. Second, the inhabitants, whether older veterans or newly arrived immigrants, were neither prepared for, nor predisposed toward, this artificially rural atmosphere. They were true urban dwellers, not rural or suburban natives. The opportunities for farming and being self-sufficient as far as certain individual family foods were concerned held no interest for these urbanites and were merely a holdover from the agrarian biases of the planners. Furthermore, in view of the fact that the development towns would

eventually grow larger, the costs of maintaining the low densities rather than the normally higher densities of urban settlements would have been prohibitive. In the mid-1950s, therefore, there was a change in planning models and a new series of more dense construction programs was begun throughout the development towns.

Another aspect of the garden city model which was inappropriate to Israeli development towns was the radial plan with its isolated neighborhoods. This plan resulted in services being dispersed throughout the towns without coordination or efficient use of mass transportation to make them available to the residents. Later development towns such as Kiryat Gat and Karmiel are based on a linear model which allows neighborhoods better access to one another and centralized, more efficiently placed services (Spiegel, 1966:61).

The development towns were conceived to absorb new immigrants, and during 1961-1967, the new towns took in 43.3 percent of all immigrants. The more startling fact is that 51 percent of all new settlers from 1957-1967 in the development towns were of Afro-Asian origin, a proportion which is higher than that of the country as a whole (Lichfield, 1970: 75). The other settlers in the new towns are either Vatikim (often induced to move to the new towns by advantages aimed at them in particular) or new immigrants of American or European origin. What seems to be a trend toward cultural homogeneity (obviously weighted to the "Oriental" Jews) masks a situation of great heterogeneity. At first, "no clear policy was evolved by the settlement authorities about the desirable ethnic composition. . . . The variation in composition . . . is mainly influenced by the period at which the bulk of immigrants were brought to a certain town as well as upon later rates of fresh arrivals and departures of immigrants. The higher the rates, the more varied the population tends to be" (Cohen, 1970a: 597). This situation has altered recently. In later towns, there are "ideal quotas" set for ethnic composition (e.g., Karmiel), but even these may not be met because of unexpected fluctuations in immigration.

The differences in occupational experiences, values, cultures, languages (of country of origin), different religious traditions, and previous exposure to members of other ethnic groups result in serious tensions among the groups. They must not only adjust to living in a new, modern, and confusing society, but they must also adjust to living with each other. Opinion is divided on whether heterogeneous neighborhoods promote friendlier living situations within the town as a whole. Shuval (1963) and Soen (1963) feel that heterogeneous neighborhoods do, in fact, lead to better relations. Erika Spiegel (1966) however, notes that "even after many years, spatial interaction among groups is at a minimum." One might easily theorize that other factors, such as the economic opportunities, the feeling of isolation in a development town, and living conditions in general, have much to do with unamicable relations among people.

The differences are exacerbated by another factor: the relations between the Vatikim and the new immigrants. There are certainly inducements for both

the Vatikim and the new immigrants to move to the new towns, e.g., investment easements, incentives for industrialization, new, low-cost housing, etc. (Pincus, 1969). These inducements, however, are aimed more specifically at the older, more experienced, and supposedly more stable Vatikim. Policy statements have resulted in special programs providing housing for new immigrants at low rents with options to buy. The rents are so low, however, that there is little incentive to buy the homes. By far the more favorable opportunities for housing (and employment) go to the Vatikim, in the form of schemes for "Saving for Housing" and "Housing for Young Couples." The housing standards and maintenance schedules in the latter programs are noticeably better than those in the immigrant housing (Spiegel, 1966).

The planners wished to capitalize on the stability which the veterans could bring to the new towns. As Erik Cohen notes: "The veterans were considered to be representative of the absorbing society towards the immigrants since they would serve both as an example and a communicator, and thus contribute to the immigrants' cultural absorption" (Cohen, 1970a: 599). It is true that the veterans, with many personal connections "in higher places" could "get things done" in the towns, but this placed the development of local leadership among the new immigrants at a disadvantage.

While traditional leaders exist [among the new immigrants] they are not directed towards development of a modern culture in a democratic society. The administrative leadership that does exist, therefore, generally comes from outside the town (Lichfield, 1970: 77).

This leadership also rests in the elite veteran population. Thus, the gap between new, mainly Oriental, immigrants and the veteran residents shows potential for growing wider. The circumstances of "dual leadership" are not irreconcilable; however, only time will tell what progress will be made.

The labor force and employment patterns depend upon the amount and kind of capital invested in the development towns and the employment potential of the residents. In the first phases of planning and building, the towns were intended to be sources for employment in village industries (agricultural or otherwise) or service outlets. The capital invested in the towns, therefore, left little possibility for an independent economic base of industry or large-scale trade located in the towns themselves. The towns, which were supposed to rely on the rural areas around them, did not grow and, perhaps, stagnated. The government accordingly has taken concrete steps to provide sound, independent economic bases for the towns (e.g., mercantile and industrial interests).

The age distribution in the development towns is younger than that of the country as a whole. This causes a situation where there are proportionally fewer able-bodied working people and a higher dependency ratio. Given the relatively lower educational and occupational preparation of many of the Oriental immigrants, they are often unable to take advantage of the jobs which do exist or are discouraged from rising on the scale of employment



positions. Furthermore, there are increasing complaints of discrimination against the Oriental Jews by the elite Westernized Jewish population.

The consequences of these factors are that although the population of the development towns has grown steadily but slowly, there is a large movement of outmigration. The population of the towns grows because of influxes of newer immigrants. The immigrants who arrived in the waves before them have become frustrated by the processes of absorption and employment (or lack thereof) and leave to find jobs and cultural accommodation elsewhere. The veterans may also leave because the new towns are not growing fast enough economically, culturally, and in standard of living. Nathaniel Lichfield has noted that even the towns which are labeled "successful" are suffering from outmigration. The term "success" denotes industrial concentration, and the citizens who leave these towns are seeking more and better human services. The difficulty of developing a core of stable, loyal residents perpetuates the transient nature of the development towns. In terms of raw numbers, the towns have achieved their purpose of dispersing the population—at least on a temporary basis—where individual families are concerned, but the development towns have yet to accomplish the task of integrating new immigrants into the fabric of Israeli society.

#### Summary and Conclusions

This paper has attempted to show that ideology provides goals (e.g., reclamation of the land, normalizing Jewish life) but insufficient means of achieving these ends. In a problem situation where reality necessitates immediate response, there may be some reliance on ideologically accepted courses of action. For example, the ideological goal of reclaiming the land coincided with the real national objective of dispersing the population. Zionist ideology suggested a course of national development which disregarded the previous patterns of settlement and characteristics of the immigrants (both predominantly urban). In the short-term, ideological predispositions may aid in solving such problems if concomitant social controls are strong enough to enforce the policy. In other words, once a mechanism for population dispersion is chosen (i.e., the development town), there must be rewards or other influences to keep settlers in the towns.

The neglect of the values of urban settlement by the ideology and by initial planning efforts provoked severe strains on the new towns by denying them a positive self-image and restricting initial economic growth (Cohen, 1970b). This myopic view also impeded development of an efficient agricultural system which could sustain a growing population with limited agricultural resources.

The hierarchical model would have maintained the small agricultural holdings of the collectives (which produced small quantities of many agricultural products) instead of lending itself to the consolidation of large-scale production of fewer products by each collective village and the emergence of

a more modern agricultural market. This latter system eventually evolved in spite of, rather than because of, the choice of the hierarchical model.

Socioeconomic and cultural characteristics of both the veteran and new immigrant groups were the decisive factors in the viability of the individual development towns. The semantic values given to the Vatikim, as a pioneer, and the Olim, as greenhorns, clearly influenced and reinforced the image and status of the towns themselves. While the early settlers in the urban centers (pre-1948) were never accorded equal value with their pioneering brothers, they, at least, had some viable skills which allowed them to participate in Israeli society. For the majority of the newer immigrants, however, who lacked both economic and social stability, the minor status given them (as Olim and as urban dwellers) was a continual reminder of their relative lack of worth in the social strata.

Thus, one must conclude that the contribution which ideology makes to the more difficult process of social integration of a diverse society is also limited. Socioeconomic characteristics are too complex for idealistic goals. Ideological principles must eventually adapt to the situation by compromising with the demands of reality. The final acceptance of urban development and urban life was a result of social and economic necessity. In a sense, the emergence of the city as a necessary tool for aiding in the country's development is an example of how ideology eventually came to be adapted to the realities of Israel's needs.

Planning and implementation are also dynamic processes, and improvements have been made in successive plans. For development towns to reach their full potential, policymakers should recognize the continuing need for (1) updating national goals and strategies; (2) instituting economic development schemes in which the villages come to depend upon the towns for trade and services; and (3) instituting social planning within the towns to respond to the particular needs of the immigrant and veteran residents.

#### References

- Akzin, B., and Y. Dror (1966) *Israel: High Pressure Planning*. Syracuse N.Y.: Syracuse University Press.
- Berler, A. (1970) *New Towns in Israel*. Jerusalem: Jerusalem University Press.
- Berger, P., and Luckmann (1967) *The Social Construction of Reality*. Garden City, N. Y.: Doubleday & Co.
- Brutzkus, E. (1964) *Physical Planning In Israel*.
- Cohen, E. (1970a) "Development Towns: The Social Dynamics of 'Planted' Urban Communities in Israel," in S. N. Eisenstadt, ed. *Integration and Development in Israel*. New York: Praeger.
- (1970b) *The City in Zionist Ideology*. Jerusalem: Institute of Urban and Regional Studies.

- Eisenstadt, S. N. (1970a) "Israeli Identity: Problems in Development of Collective Identity," in S. N. Eisenstadt, ed. *Integration and Development in Israel*. New York: Praeger.
- (1970b) "Israel: Traditional and Modern Social Values and Economic Development," in S. N. Eisenstadt, ed. *Integration and Development in Israel*. New York: Praeger.
- Fein, L.J. (1967) *Politics in Israel*. Boston: Little, Brown.
- Gouldman, M. D. (1966) *Legal Aspects of Town Planning in Israel*. Jerusalem: Institute for Legislative Research and Comparative Law.
- Lichfield, N. (1970) *Israel's New Towns: A Strategy for Their Future*, Vol. I of Stage I Draft for Discussion. Tel Aviv: Institute for Planning and Development.
- Manpower in Development Towns: First Report on Twenty-One Development Towns* (1964, Dec.) State of Israel, Ministry of Labour, Man Power Planning Authority.
- Matras, J. (1965) *Social Change in Israel*. Chicago: Aldine.
- Princus, L. A. (1969) "The Impact of Aliya on Israel's Development," page 8 in L. Berger, ed. *The Impact of Aliya on Israel's Development*. Jerusalem: Jewish Agency.
- Shachar, A. (1971) "Israel's Development Towns, Evaluation of National Urbanization Policy," *Journal of the American Institute of Planners* 37, no. 6 (Nov.)
- Shuval, J. (1963) *Immigrants on the Threshold*. New York: Atherton.
- Spiegel, E. (1966) *Neue Städte/New Towns in Israel: Urban and Regional Planning and Development*. Stuttgart: Bindearbeiter C. H. Schwabe.
- Strong, A. L. (1971) *Planned Urban Environments: Sweden, Finland, Israel, The Netherlands, France*. Baltimore: Johns Hopkins.
- Weingrod, A. (1965) *Israel: Group Relations in a New Society*. New York and London: Praeger.

---

# 32 Administrative Pluralism and Italian Regional Planning

Robert C. Fried

Analysis of regional planning in Italy is difficult due to the fact that the phenomenon is both multiform and amorphous.\* There has not been any clearly individuated, discrete phenomenon, but rather a diversity of forms, institutions, practices and intentions. There have been several competing forms of regional planning; several competing institutional sponsors of regional planning; several competing concepts of regional planning; and several competing sets of regional planners. Some of the competition has come from *proposed* forms, *proposed* concepts, from *potential* planners. Indeed, the second quality – amorphousness – stems from the fact that much of Italian regional planning is intentional rather than actual, and is composed of proposed schemes, often in the form of draft legislation, or as yet unimplemented constitutional provisions, rather than approved, on-going institutional practice. Post-war Italian regional planning, then, consists of several inconclusive experiments, plus a rather ambitious but constantly evolving set of proposals for the future, involving major institutional innovation. The interest of Italian regional planning lies partly in the multiformity of past experiments and partly in the grandiose architectonics of proposals for the future.<sup>1</sup>

## ORIGINS OF ITALIAN REGIONAL PLANNING

The multiformity of postwar Italian regional planning derives from the fact that it resulted from the confluence of several distinct strands of

\*Material for this article was gathered in Italy during the period August 1966-February 1967. Assistance from the Economic Development Administration (U.S. Dept. of Commerce) and the Academic Senate of the University of California, Los Angeles, is gratefully acknowledged. I would also like to thank the following for their help: the Centro Italiano Ricerche Documentazione (Rome); Dr. Alfredo Testi (Ministry of the Budget); Ing. Mario D'Erme (Cassa per il Mezzogiorno); Dr. Luigi Pierracioni (Union of Italian Chambers of Commerce); Dr. Domenico Sartoni (Union of Italian Provinces); Prof. Gabriele Scimemi

political, institutional, ideological and economic development. Its origins can be traced back to a half dozen distinct socio-political movements: (1) the movement to develop the Italian South; (2) the movement to establish semi-autonomous regional governments; (3) the movement to establish effective urban planning; (4) the movement to establish national economic planning; (5) the movement towards European integration; and (6) the movement to reform Italian public administration. Each of these reform movements had its own goals, organization, traditions and interests. Each of them eventually came to see in regional planning a necessary solution to some of the problems it considered important.

(1) The overriding problem of postwar Italian economic policy, apart from stability, has been the differential social and economic development of the North and the South. While northern and central Italy developed a modern society and economy similar to that of the rest of western Europe, southern Italy remained frozen in a semi-feudal mold of backwardness. The disparity between North and South created and sustained the interest in comparative regional development that has lain behind other experiments in regional planning. Action began on the problem in 1950 with the establishment of the Fund for Southern Development (*Cassa per il Mezzogiorno*) and the commencement of the land reform programme. Control of the programme was centralized in Rome and the programme was carried out on a super-regional scale, covering the whole of the South plus the islands of Sicily and Sardinia. Efforts to reduce the gap between North and South were frustrated by the economic boom of the late 1950's, which, by increasing the gap between the sections, brought the question of uneven territorial development to the top of the policy agenda. As southern development lagged, especially in the industrial sector, the Southern Development Fund itself began to experiment in decentralized regional planning, in 1957, in the form of 'consortia for industrial areas' (*consorzi per le aree industriali*).<sup>2</sup>

(2) The movement to establish semi-autonomous regional governments derived from a reaction to the extreme centralization of government under the Fascist regime, when all local authorities were appointed from above. After the fall of the Fascist regime the decentralization movement managed to insert into the new constitution, promulgated on 1 January 1948, provisions for the establishment of semi-autonomous regional governments to operate at a level between the national government in Rome and the ninety provincial and approximately 8,000 municipal governments (or communes) that were the traditional pre-Fascist forms of local government.<sup>3</sup> There were to be five Special Regions (Sicily, Sardinia, Trentino-Alto Adige, Aosta Valley, and Friuli-Venetia Julia) and fourteen Ordinary Regions (Lombardy, Venetia, Piedmont, Emilia-Romagna, the Marches, Tuscany, Umbria, Liguria, Latium, Abruzzi-Molise, Campania, Apulia, (Centro di Studi e Documentazione per la Pianificazione Territoriale, Istituto di Urbanistica, Fac. di Architettura, Univ. of Rome); and Dottorssa Gloria Pirzio Ammassari.

Lucania, Calabria. In 1965 the Region of Abruzzi-Molise was split into two separate regions.) The Special Regions were to be given greater autonomy and greater governmental authority than the Ordinary Regions, either because they were islands (Sicily and Sardinia) or because they contained large ethnic minorities (the French-speakers in Aosta Valley; the German-speakers in Trentino-Alto Adige; the Slav-speakers in Friuli-Venetia Julia). For various reasons or pretexts – including the threat of possible Communist control in the regions of central Italy, the disinclination of national functionaries to become regional functionaries, the not always edifying performance of some of the Special Regions, the possible costs of operating fifteen new governments, the resentment and envy of the large number of cities and towns *not* chosen to become one of the fifteen regional capitals, and the general reluctance to carry out major institutional change by the ruling centre and centre-right party coalition – the Ordinary Regions have not yet been established. Almost every cabinet since 1948 pledged itself to establish the Ordinary Regions. But only on 14 February 1968 did this pledge finally become a legally binding commitment. On that date, the Italian Parliament, after protracted debate, approved a law calling for the election of the fifteen regional assemblies at the local government elections of 1969.

In most parts of Italy, therefore, there have been no locally elected and responsible authorities at the regional level to carry out regional planning. The Special Regions, however, have been created – due in most cases to the separatistic or autonomistic drives in those areas – and they have been active in producing schemes for regional economic development. The Ordinary Regions, when created, will not have authority to make regional economic plans, but only physical development plans: when the regions were designed, in fact, there was little thought about their possible role, or the role of the national government, for that matter, in economic development.<sup>4</sup>

(3) Physical planning (i.e. town and country planning) in Italy is regulated by the Urban Planning Act of 1942 (the Law of 17 August 1942, No. 1150) – the so-called *Legge Urbanistica*, considered to be one of the better pieces of legislation produced by the Fascist regime. This act regulates and authorizes physical planning at the municipal (communal), inter-municipal ('intercommunal'), and 'territorial' levels. Plans developed by municipalities singly or jointly require the approval of the national Ministry of Public Works. 'Territorial' plans (called that, rather than regional plans because of the ideological aversion of Fascism to 'regionalism' as both word and concept) were to be made directly by regional field offices of the Ministry.<sup>5</sup> The 1942 Act did not specify what the boundaries of the 'territories' were to be. When the Ministry in 1950 began to create 'territorial plans of co-ordination', as they were called in the Act, it used the boundaries of the nineteen Special and Ordinary Regions, which in turn derived from groupings of provinces used for statistical purposes in

censuses after unification. Thus the constitutional regions were to coincide with the physical planning regions. When and if the Ordinary Regions were established, they would take over control of regional (physical) planning from the regional field offices of the Ministry of Public Works. Italian city planners turned increasingly to regional planning as the inadequacies and frustrations of planning at the lower levels became increasingly evident: the problems of cities and metropolitan areas could not be solved, they came to feel, within the narrow framework of municipal boundaries. The national organization of Italian city planners – the *Istituto Nazionale di Urbanistica* – held a national conference on regional planning as early as 1952.<sup>6</sup> By the late 'fifties, however, the city planning movement began to press for a new general planning law to replace the Act of 1942 and to formulate new schemes of regional, 'district', and local planning closely tied to the then emerging process of national economic planning. This reflected a growing belief on the part of physical planners that their plans were futile if not co-ordinated with and reinforced by national and regional economic planning decisions.<sup>7</sup>

(4) The 'economic miracle' of the late 1950's, while producing a great leap forward in the living standards of urban and North-Central Italy, also produced a deep concern with the failure of southern and rural areas to keep pace with the general advance. Massive exodus from the South and the countryside into a few metropolitan areas – notably Rome, Turin, Milan and Genoa – also forced attention to the problems of congestion and overdevelopment in the favoured areas. As Italian economists and policy-makers became more development-oriented, they also began to ponder the problems of territorial imbalances and the adverse physical consequences of economic development. In 1954, Minister of the Budget, Vanoni, a Christian Democrat, had produced a Ten-Year Development Scheme, involving more a set of predictions than of policy directives. In 1959, Christian Democrat Minister of Industry and Commerce, Colombo, faced with a flood of demands for subsidies and 'incentives' from all parts of the country, established a series of regional planning committees – which became known as 'Colombo Committees' – in the chambers of commerce of the regional capitals.<sup>8</sup> (Italian Chambers of Commerce are semi-official agencies of the national government, which appoints their presidents.<sup>9</sup>) These committees were to help the Ministry decide among competing claims for state aid to industry and to help verify the predictions of the Vanoni scheme. Three years later, in 1962, Italian economic policy shifted somewhat to the Left, with the formation of the alliance between the Christian Democrats and the Socialists, on the basis of inter-party agreements calling for the production of a national economic plan, as well as a new urban planning Act. The Ministry of the Budget, responsible for preparing and executing the national economic plan, established in 1964 another set of regional planning committees to help in the formation of the national economic plan and to produce regional economic development

plans on their own.<sup>10</sup>

(5) Regional planning in Italy received further impetus from active Italian participation in the European integration movement, including the Council of European Communes, the European Federalist organization, and the organs of the European Economic Community. Italian representatives of these organs helped develop a European point of view on regional planning, heavily emphasizing the need for *democratic* regional planning, which in turn became one of the sets of standards by which the Italians judged the quality and performance of their own institutions. Art. 2 of the Treaty of Rome (1957) which gave rise to the European Common Market placed on each member nation the obligation 'to reinforce the unity, and to ensure the harmonious development of their economies, reducing the disparities between the various regions and the retardation of those less favoured'. European integration, by juxtaposing rich and poor nations, led to the same kind of concern with uneven development produced by the juxtaposition of North and South within Italy itself.<sup>11</sup>

(6) Regional planning was also affected by the general crisis in Italian public institutions, overwhelmed by the growing demands and complexities of a rapidly modernizing society. Economic growth and accompanying massive population movements placed great strains on the national and especially local governments. There was a growing gap between the felt needs of Italian society and the capabilities of Italian public administration. Regional planning came to be seen as a technique for improving the quality and effectiveness of government action, stimulating the national ministries to step up and co-ordinate their activities and providing a new *raison d'être* for local institutions – municipalities, provinces, chambers of commerce, and existing and future regional governments. Regional planning would provide an incentive for local initiative and allow such initiative to compensate for the widely recognized disabilities of the national government.<sup>12</sup>

From these diverse streams emerged several distinct experiments in regional planning: (1) the Industrial Area Consortia (1957-present), sponsored by the Southern Development Fund; (2) the regional development planning of the Special Regions, particularly Sicily and Sardinia; (3) the physical planning efforts of the Ministry of Public Works and its 'territorial plans of co-ordination'; (4) the regional economic planning activities of the Ministry of Industry and Commerce (1959-62); (5) those of the Ministry of the Budget (1964-present); and (6) the development efforts of groups of local elected authorities, particularly the provinces.

#### SUCCESSIVE EXPERIMENTS IN REGIONAL PLANNING

##### (1) *The Industrial Area Consortia (Southern Development Fund)*

The programme to develop the Italian South has generally been conducted as a rather centralized operation on a super-regional scale. In 1957 an



effort was made to make the programme more specific and more effective by concentrating development efforts in a selected number of zones and by requiring and stimulating the collaboration of local elected governments in development programmes, particularly industrialization. The law of 29 July 1957, No. 634, called for the establishment of Industrial Areas Consortia (*Consorti per le aree industriali*) on the initiative of local governments in those areas in the South and Islands with particularly favourable industrialization potential. Minimum requirements in population, existing industry, municipal participation, and physical, social and economic characteristics were specified before a Consortia could be recognized as a public law institution, with the right to request several forms of aid from the Southern Development Fund, and with the obligation to draft a development plan for the area and to manage the collective services created under the plan.<sup>13</sup>

Despite the specification of minimum requirements and the declared policy of concentrating industrialization efforts, the number of Consortia (and lesser bodies called Nuclei) multiplied to such an extent that practically the whole territory of the South and Islands is now covered. It has been politically impossible to exclude most localities from inclusion in industrialization efforts, even though this means dispersion of resources and effort. Further difficulties have stemmed from the inability of large numbers of southern local governments to make any significant financial contribution to the Consortia. As a result, most Consortia have been almost completely dependent on aid from the Southern Development Fund. Local initiative has often meant, at best, initiative in requesting central aid. It has not been easy to make southern local governments into active participants in development efforts.<sup>14</sup> For this reason, schemes of decentralized regional planning and development are greatly feared by the South and by those concerned with southern development: decentralization threatens to allow the wealthier, more dynamic regions to become even more so, and to leave the less developed regions ever further behind. Regional planning, it is feared, may become, not a means of overcoming unbalanced development, but instead the unconscious instrument for aggravating inequalities.<sup>15</sup>

## (2) *Planning in the Special Regions (Sicily and Sardinia)*

Of all the Special Regions created, Sicily was granted the broadest amount of governmental authority. Article 14 of the Statute of the Sicilian Region, approved by the Italian Constituent Assembly in 1948, gave exclusive authority to legislate on physical planning in Sicily to the Sicilian Regional Assembly. In October 1952 the Regional Government submitted a bill on physical planning to the Regional Assembly, where it has since remained; consequently, the national Urban Planning Act of 1942 is still valid on the island, with some exceptions. In 1953, the Regional President created

a Regional Planning Commission (*Consiglio Regionale Urbanistico*) with authority not only to approve the master plans of the Sicilian municipalities but also to draft a regional physical plan. The regional plan has never been produced. The Commission still exists, with powers to approve city plans, but it shares these powers with the regional field office of the national Ministry of Public Works. Failure to produce a Sicilian planning law has left authority over physical development divided and uncertain as between regional and national government officials.<sup>16</sup> The results have in some cases been literally disastrous, as in the Agrigento landslide of 1965, when many lives were lost due to failure to regulate building and planning practices in the city.<sup>17</sup>

Article 38 of the Sicilian Statute provided for an annual national grant to the Sicilian Region, constituting a 'Fund of National Solidarity', to help compensate for the disparity between Sicilian and 'continental' Italian income. The annual grant was to be spent on public works programmes, which would be parts of a long-range development programme. In 1956, the Regional President appointed a committee to produce a Five-Year Plan for the Economic Development of Sicily. In four months a plan was produced (the so-called 'Aldisio Plan', 15 May 1956) calling for greatly increased investment especially in agriculture by the national government, the Fund for Southern Development, and the Regional Government. This plan does not seem to have ever been implemented due to (1) the instability of the Regional executive, a frequent victim of cabinet crisis; and (2) the disinclination of regional political and administrative forces to restrict their freedom of manoeuvre and action in accordance with a pre-ordained allocation of resources.<sup>18</sup>

A second economic plan was formulated in 1964-65. The prospects for this plan seem to depend on whether sufficient public interest can be aroused so as to make it dangerous or at least disadvantageous to disregard the commitments it contains.<sup>19</sup>

The Sardinian experience with regional self-government has been less dramatic, as well as less controversial than the Sicilian. Sardinia received less home rule in its 1948 Statute than Sicily, and the national government and its agencies, such as the Southern Development Fund, have remained much more closely involved in Sardinian affairs than in Sicilian affairs. Article 13 of the Sardinian Statute, approved in 1948, reads: 'The State [i.e. the national government], with the assistance of the Region, shall prepare an organic plan to promote the social and economic rebirth of the Island.' The Committee of Ministers for the South – the cabinet committee in the national government in charge of the Southern Development Fund – in 1951 established, jointly with the Regional Government, an advisory commission to make a survey of Sardinian resources and formulate a development programme. Lack of funds delayed presentation of the commission's report until 1958, when it was presented to the Chairman of the Committee of Ministers for the South. To implement the report, the

Region established a development department ('Rebirth Assessorate') while the Committee of Ministers, jointly with the Region, established a Working Group to prepare an action programme. This programme was presented to Prime Minister Segni, the leading Sardinian political leader, in 1959. After being drastically reduced in cost and scope, it became the Law of 11 June 1962, No. 588. Responsibility for implementing the 1962 Act was given not to the Regional Government, as requested by island authorities, but to a Special Section of the Southern Development Fund. The Region was given responsibility for presenting proposals on the long-term plan and annual programme, but final word rested with the Committee of Ministers for the South. The law authorized a lump-sum appropriation of 400 billion lire (about \$667 m.) to be spent in thirteen fiscal years (1962-1975). In theory, this grant was to be additional to the regular spending by national agencies on the island, but the law provided no guarantee that these agencies would not, as they have regularly tended to do, cut down their spending in areas, such as the South and Islands, that benefit from special aid legislation. Nor did the law provide any guarantee that the various State and regional agencies would act in accordance with the objectives and directives of the Rebirth Plan.<sup>20</sup> Nonetheless, unlike Sicily, Sardinia has actually formulated and funded a regional development plan and achieved for its own Regional Government a large, if not dominant role in the regional planning process.

### (3) *The 'Territorial Co-ordination Plans' (Ministry of Public Works)*

The Ministry of Public Works began in 1950 to implement Article Five of the 1942 Planning Act, which provides for the formulation of 'territorial co-ordination plans' for the purpose 'of orienting or co-ordinating urbanistic<sup>21</sup> activity in determinate parts of the national territory'. The plans were to indicate (1) 'zones to be reserved to special uses and those subject to special restrictions or legal limitations'; (2) 'localities to be chosen as sites for new building nuclei or installations of a particular nature or importance'; and (3) 'the network of the principle lines of highway, rail, electric, and navigable communications, both existing and planned'. They were to be drafted 'in agreement with [*d'intesa con*] the interested administrations' and approved by decree of the chief of state on the joint proposal of the Minister for Public Works, the Ministry of Communications (now Transport) for railway lines, and the Ministry of Corporations (now Industry and Commerce), for the location of industrial zones. As can be seen, the Act was very vague with regard to what the plans should do or contain. The effectiveness of these provisions of the 1942 Act was further undermined by the provision of the 1948 Constitution which delegated the whole matter of *urbanistica* or physical planning to the Regional governments, whenever they might be set up. The Ministry nonetheless estab-

lished advisory committees of state functionaries and outside experts in some of the regional field offices of the Ministry to prepare regional plans under the supervision and control of the regional field directors of the Ministry. These committees proceeded to collect information about regional resources in accordance with a standard inventory prepared by leading Italian planners and published by the Ministry in 1952-53.<sup>22</sup>

Despite the provision of a standard inventory, data collection proceeded in rather different fashions in the various regions, making inter-regional comparisons very difficult. In some cases, after initial and unsuccessful attempts to secure data, the whole task was abandoned. In others (Campania, Abruzzo, Emilia-Romagna, Lombardy), an imposing amount of information was collected and analysed, providing very useful surveys of regional social and economic life. In only a few cases, however, such as Lombardy, was data collection followed by an attempt to define the physical and economic traits that each of the several parts of the region should keep or acquire in the course of development. But even in these cases, the studies remained without operational consequences.

This was due to several things. (1) The studies were not tied to a process of national and regional economic planning: they attempted to plan infrastructures without relating their decisions to those of public and private investors. Economic factors were taken to be either not subject to control or as independent and not necessarily relevant. (2) The plans were prepared for institutions that had neither the desire nor the intention to use them, and often not even willingness to supply basic data; and which preferred to retain a completely free hand in their investment decisions. Public agencies, in particular, were anxious to remain visible and identifiable as the source of good deeds in the region, rather than merge their personality in an impersonal collective planning operation. (3) The plans had no binding effects either on such agencies or private investors. (4) The Ministry itself was only intermittently interested in the fate of the plans, which often lay unnoticed in the Ministry for several years before they were published. By this time, the data upon which the plans were based were out of date. The Ministry, in any event, was not prepared to fight all the different agencies creating infrastructures in order to secure their commitment and adherence to its 'territorial co-ordination plans'. (5) Local elected officials were either excluded from participation on the regional committees, or given a strictly marginal role so that their self-interest, and that of the governments they represented, was not harnessed to the planning operation. Greater local participation might have made the resulting plans less 'rational', but more urgent and less easily ignored.<sup>23</sup>

After a decade of relative inactivity as far as regional planning was concerned and after the entry into the regional planning field of competing agencies, notably the Ministry of Budget, the Ministry decided in 1965 to relaunch its campaign to formulate regional physical plans under the authority of Article Five of the 1942 Planning Act.

(4) *Regional Economic Planning: I (Ministry of Industry and Commerce)*

In 1959 the Ministry of Industry and Commerce called for the establishment in the chambers of commerce of the regional capitals of Regional Economic Development Committees. Such committees were set up in several regions in the following three years. The presidents of the chambers of commerce of the regional capital became the chairmen of the regional planning committees. The staffs of the chambers of commerce became the committee staffs. These instrumentalities were chosen because, among other reasons, the chambers were and are the field organs of the Ministry of Industry and Commerce, which appoints and can remove their presidents. The committees included the presidents of the other chambers of commerce in the region; the presidents of the provinces contained in the region; the regional field directors of the Ministries of Public Works, Agriculture, and Labour; representatives of management and labour; the presidents of the provincial tourist agencies; and representatives of the state railways, land reform agencies, reclamation 'consortia', etc. There were no representatives of the municipalities, and national functionaries had an overwhelming preponderance of representation. Each committee was assisted by a working group of no more than eleven experts in economic development, appointed by the chairman of the Regional Committee. The purpose of the committees was to make surveys and calculate trends and development potentialities so as to make it possible for the Minister of Industry to make more equitable and rational allocations of aid to industry. (They also served to demonstrate the acceptance by moderately conservative Christian Democratic circles of at least some form of economic planning.)<sup>24</sup>

For various reasons, the Regional Economic Development Committees – the first to focus on the economic, rather than the physical aspects of regional development – never became vital or productive institutions. Most of the committees lapsed into complete inactivity and contributed little to the developing theory or practice of regional planning in Italy. For one thing, the presidents of many chambers of commerce did not believe in the legitimacy of economic planning. For another, many chambers of commerce employed no economists trained in modern economic theory and capable of making useful studies in development.<sup>25</sup> (The Italian Union of Chambers of Commerce in Rome began to sponsor graduate courses in economic development to meet this deficiency.) Also, exclusion of representatives of the municipalities from the committees deprived them of a source of energy which many of them apparently needed. Local politicians might have been more interested in stimulating development through public action than local business elements evidently were.<sup>26</sup>

(5) *Regional Economic Planning: II (The Ministry of the Budget)*

The latest and most significant experiment in regional planning was begun by the Ministry of the Budget in 1964. Its decree of 22 September

1964 called for the establishment in all the non-Special Regions of Regional Committees for Economic Planning. The new committees included the president of each province (there are about four provinces in each region); the mayors of each provincial capital and each municipality (commune) with over 30,000 people; three economic experts appointed by the Ministry of the Budget; a representative of the Southern Development Fund in the areas where it operates; the presidents of the chambers of commerce (one from each province); the regional field directors of the Ministry of Public Works and the Ministry of Agriculture; three representatives each of labour and management; plus a chairman appointed by the Budget Ministry.<sup>27</sup>

The title, composition, and sponsorship of this latest venture in regional planning reflect the shift in Italian politics that took place in 1962, when the previous centre or centre-right coalitions gave way to a centre-left coalition, running from the Christian Democrats in the centre (who have dominated every cabinet since the war) to the Socialists on the left (but excluding the Communists). Socialist participation in the government was arranged on the basis of inter-party agreements pledging the enactment of a national economic plan, the establishment of the Ordinary Regions, the passage of a new urban planning Act, the nationalization of the electrical industry, as well as other less spectacular reforms. A Socialist Minister of the Budget proceeded to establish the Regional Economic Planning Committees, which differed in several respects from the Regional Economic Development Committees or Colombo Committees established by the Christian Democratic Minister of Industry and Commerce in 1959-62 – before the shift to the left. The title change indicated a more explicit commitment to planning as opposed to mere prediction. The new committees gave sizeable representation to the municipalities, excluded from the Colombo committees; the representation of labour and management was equalized to three and three, instead of the former five to three in favour of management; most of the state field offices represented on the Colombo committees were no longer represented; three experts appointed by the Minister of the Budget were now committee members, as compared to the experts appointed by the Colombo committee chairman who remained outside advisers. The new committees are located in the prefectures (i.e. the offices of the local representatives of the Ministry of the Interior, in charge of the police, supervision of local government, and general representation of the national government); the Colombo committees were located in the chambers of commerce.

The functions of the Colombo committees were to 'study the prospects of development of the economy of the provinces, including the evolution of employment' (Decree of 8 September 1961). The functions of the Regional Economic Planning Committee have been defined more precisely in the decree of 22 September 1964. They are to assist the Ministry of the Budget in the administration of national economic planning until such

time as the regional governments are established; at that time, the regional governments will become what the decree calls the 'natural interlocutors of the Government for the purpose of national economic planning'. For this reason, the Ministry relies on the regional governments where they already exist (i.e. in the five Special Regions) and on the Regional Economic Planning Committees, where regional governments have not yet been established (i.e. in the fifteen Ordinary Regions). The role of the Regional Planning Committees is to identify local economic resources; the problems of economic development; and the possible objectives and means of intervention in the region. They are then to 'prepare, in accordance with the directives of the Ministry of the Budget, a draft plan for regional economic development'. And they are to supply the Ministry 'with all information the latter requires relating to the activity of economic planning'.<sup>28</sup>

The initial reaction of the Left to the establishment of these committees was favourable, while the reaction of the conservatives was not. Gradually the positions have become reversed. The conservatives have accepted the existence of the committees and made considerable efforts to secure a large, if not dominating position on them.<sup>29</sup> The Left approved the initial inclusion of municipal representatives; the deference to the existing and future regional governments; the equalization of labour and management representation; and the definite assignment of planning functions.<sup>30</sup> (In Italy, the Left since 1947 has been the strongest defender of grass roots power, while the Centre and Right have defended the prerogatives of the central government. This is not unrelated to the fact that the Left from 1947 to 1962 has been excluded from power at the national, but not at the local level. It has therefore favoured all measures which would give local authorities greater power. The same kind of reasoning, reversed, has kept the Centre and Right from creating the Regions and increasing the power and autonomy of the provinces and communes.)

But the Left has become rather critical of the new committees, asserting that (1) they are dominated by the Ministry of the Budget and do not really allow full participation by local interests in the economic planning process; (2) they are excessively large and heterogeneous combinations of politicians, bureaucrats, group representatives and experts, among whom it will be very difficult to develop a common language or purpose.<sup>31</sup> The Ministry, in responding, is caught between the conflicting requirements of two basic ideals of Italian reformism: economic planning vs. representative, grass-roots democracy. Economic planning seems to require strong centralized direction, and for three major reasons: (1) the need for uniformity of methodology and approach among the regions so as to arrive at quantitative formulations that can be interrelated; (2) the need for central impulse and direction in the less developed areas; and (3) the need to inject elements of realism into calculations of regional needs and development potentials. Representative, grass-roots democracy seems to require that as many local elected authorities as possible participate in the formulation and execution

of plans affecting their constituencies – and that they be given sufficient discretion to make their participation meaningful and worthwhile. The structure of the Regional Economic Planning Committees represents an attempt to satisfy both ideals – but an attempt which seems to favour central technocratic logic more than the logic of effective local participation.

(6) *Economic Planning by Local Governments*

Italian local governments – municipalities and provinces – have not been entirely absent from the regional planning scene. Indeed, some of the most important regional planning initiatives have come not from the central government but from the local authorities, especially the provinces, and especially those in the Centre and North.<sup>32</sup> Particularly important have been the regional economic research institutes created in several regions by the provinces alone or with the co-operation of municipalities, banks, private businesses, and chambers of commerce.<sup>33</sup> The national government, or at least the Ministry of the Interior, charged with the supervision of local government, attempted for some years to discourage such initiatives by provinces and municipalities, especially if under leftist control. The Ministry issued a circular in 1961 denying the legal right of the provinces to formulate or adopt development plans, or to create research institutions: these tasks, the circular said, belonged to the chambers of commerce. The circular met with a strong reaction from the provinces and the Left (including the Christian Democratic Left) and was generally ignored by the Ministry's agents in the provinces, the Prefects.<sup>34</sup> Nonetheless, there began a long conflict between the provinces and the chambers of commerce over the right to sponsor and participate in development initiatives – a conflict that has continued to the present. A Ministry of the Budget decree (15 November 1965) on the organization of the Regional Economic Planning Committees provided that these committees might confer exclusive rights to act as research staff for the committees on a regional research institute (i.e. the offspring of the provinces). With the appearance of the Regional Planning Committees, and the demise of the Colombo committees, the competition to control regional planning became a fight to control the Regional Planning Committees; this in turn has become a fight to control the technical staffs of those committees. It is sometimes suspected that these technical staffs may become the dominant element in the regional planning system – within the framework of control set by the technicians in the Ministry of the Budget.<sup>35</sup> For this reason, the chambers of commerce have conducted a vigorous campaign to prevent the assignment of exclusive research staff rights to the institutes sponsored by their rivals. The Ministry of Industry has helped in this campaign with a circular of 13 June 1966, calling for the formation in all regions of economic research centres attached to the chambers of commerce, capable of acting as the research staff of the Regional Planning Committees, especially in those regions in which regional research institutes have not yet been formed.<sup>36</sup>



There has been competition among various interest groups and political forces to secure representation or greater representation on the committees themselves. The Ministry of the Budget has been able only partially to withstand the pressure. Its decree of 7 April 1965, brought on to the committees a representative of (1) the artisans; (2) the co-operative societies; (3) the provincial tourist bureaus; (4) the land reform agencies (now called 'development agencies') in those areas where they exist. It also provided for *two* representatives of the small farmers (*coltivatori diretti*) and raised the representation of labour from three seats to four. (Raising the representation of labour from three to four was a very curious act on the part of a Socialist, who was thus providing representation on the Regional Planning Committees of the labour unions under Neo-Fascist control.) A decree of 23 April 1965 which raised the number of experts on the committee to be appointed by the Ministry from three to four, was reportedly designed to allow each of the four parties in the national government coalition to appoint a friendly expert on each of the Regional Committees.

#### CONCLUSION

It is too early to be able to tell what the outcome of these various competitions and conflicts will be, or indeed to have a clear idea of what the conflict and competition is all about. It is tempting, especially in these concluding remarks, to argue that the struggle over regional planning is a fundamental one in Italian politics and that on the outcome hinges the fate of Italian constitutional democracy. Actually the struggle is of middle-range importance and is apt to be determined by the outcome of far more significant, if closely related struggles – those involving the fate of (1) the regional government scheme; (2) national economic planning; and (3) new physical planning legislation.

The fate of Italian regional planning is tied to the fate of these three heavy-weight issues, which for some time have dominated the centre of the Italian policy-making stage, causing one dramatic political event after another: cabinet crises, individual ministerial crises, inter-party battles, as well as intra-party factional disputes. Since 1962, Italy has been governed by cabinets based on Socialist support and committed to reform; committed, that is, to the enactment of legislation on the regions, on economic planning, and on physical planning. So far none of these commitments has been met because of (1) the onset of a temporary economic recession and the need to restore 'confidence'; (2) the continued political strength of the conservative forces in and out of the government coalition who are opposed to the reforms in question 'for the time being'; and (4) the violently controversial nature of the reforms in question, involving as they do potentially major shifts in property and power relations within the country.

The final year of the IV Legislature (1963-1968) witnessed some steps forward. The Five-Year Plan was given formal statutory sanction on 25

July 1967. The 1942 Planning Act was partially updated in the so-called 'bridge-law' (*legge-ponte*) of 6 August 1967, No. 765. The Election Law for the regions was passed on 14 February 1968. But the efficacy of the Five-Year Plan has yet to be tested. The general reform of physical planning legislation has yet to be enacted. The Ordinary Regions have yet to be elected.<sup>37</sup> And, following the general elections of May 1968, the continued viability of the Catholic-Socialist reform alliance has yet to be demonstrated.

Thus the national battles over regionalism and planning are far from over. The outcome of these battles will be decisive in giving meaning and shape, if any, to the ongoing experiments in regional planning. Current regional planning ventures will either become elements in highly elaborate, finely articulated, and broadly focused national programmes in regional self-government and physical-economic planning – that is parts of a new and rationalized institutional framework – or they will share the fate of previous ventures and become more or less successful contenders in a mere struggle for institutional prestige and survival.

The irony of regional planning in Italy lies in the fact that the goal of co-ordinating planning and development efforts has led to the proliferation of would-be co-ordinators: the particularism which is the enemy of regional planning has been reproduced in the ranks of the regional planners themselves. Perhaps the new institutional arrangements, when and if they are made, will help to provide some co-ordination among the co-ordinators.<sup>38</sup>

# REFERENCES

1. For a general historical survey until 1961 of Italian regional planning, see Giacomo Corna-Pellegrini, 'Esperienze di studi e pianificazione regionale in Italia', in *Gli squilibri regionali e l'articolazione dell'intervento pubblico* (Milan: Lerici, 1962), pp. 415-502. See also the earlier survey by Alessandro Molinari, 'Metodi e criteri per l'elaborazione dei piani regionali di sviluppo', *Il Nuovo Osservatore*, 3 October, 1959, pp. 3-10, and Mario D'Erme, 'La pianificazione regionale: una via alla conoscenza e all'impegno', *Questitalia*, May-June 1961, pp. 50-58.
2. The literature on southern development is vast. The latest references may be found in the recent volumes of the *Annuario Politico Italiano*.
3. The Sicilian Special Regional already existed under the terms of Royal Legislative Decree of 15 May 1946, No. 455. The Friuli-Venetia Julia Special Region was not created until passage of the Constitutional Law of 31 January 1963, No. 1.
4. The *Annuario Politico Italiano* also contains the latest Italian references on the subject of regionalism. A brief treatment is contained in R. Fried, *The Italian Prefects*, New Haven: Yale University Press, 1963, Ch. Five. On the performance of the Special Regions, see Claude Palazzoli, *Les régions italiennes*, Paris: Pichon, Durand, 1966.
5. The 1942 Act is explicated in Virgilio Testa, *Disciplina Urbanistica*, Milan: Giuffrè, 1961.
6. See *La Pianificazione regionale*, n.p., Istituto Nazionale di Urbanistica, 1953, which contains the proceedings of the 1952 congress.
7. Fiorentino Sullo, *Lo scandalo urbanistico*, Florence: Vallecchi, 1964.
8. Emilio Colombo, *Indirizzi di politica industriale*, Rome: Unione Italiana delle Camere di Commercio, Industria ed Agricoltura, 1959.
9. Roberto Bracco, 'Le Camere di Commercio nelle situazione di oggi e di domani', *Arti e Mercature*, July-August 1964, pp. 1-4, and Unione Italiana Camere Commercio Industria Agricoltura, *The Chambers of Commerce, Industry and Agriculture in Italy*, Rome: UICCA, 1965.

10. See Joseph La Palombara, *Italy: The Politics of Planning*, Syracuse: Syracuse University Press, 1966, for the development of national economic planning in Italy.
11. See the January 1966 special issue on regional planning of *Comuni d'Europa*, also published as *La Pianificazione democratica del territorio e il problema delle regioni in Europa*, Rome: Comuni d'Europa, 1966.
12. For the viewpoint of the provinces, for example, see Unione delle Province d'Italia, *XXI Assemblea Generale Ordinaria dell'Unione: Palermo, 10-13 Novembre 1963. Atti*, and, by the same, *Assemblea Straordinaria dei Presidenti di Provincia. Sorrento, 3-4 December 1965*. See also 'Le province e lo sviluppo economico', *Questitalia*, March 1962, pp. 38-40.
13. Comitato dei Ministri per il Mezzogiorno, *Statuto Tipo dei Consorzi per le Aree di Sviluppo Industriale nel Mezzogiorno*, Rome: December 1959; Pietro Martinez Cabrera, 'Sviluppo agricolo e industriale e piani comprensoriali', *Esperienze Administrative*, June 1964, pp. 56-60.
14. Domenico Manna, 'Aspetti territoriali della politica di sviluppo industriale', in *Mezzogiorno e politica di piano*, Bari: Laterza, 1964, pp. 275-315; Mariano d'Antonio, 'La crisi dei consorzi industriali: dalla politica dei poli alla programmazione regionale', *Cronache Meridionali*, No. 11-12, 1962, pp. 17-64; Enzo Scotti, 'Lo sviluppo industriale', in *Mezzogiorno e politica di piano* (cited above), pp. 143-193 at 169-175; Diego del Rio, 'Compiti degli Enti Locali e piani di sviluppo nelle regioni meridionali', *Comune Democratico*, April 1962, pp. 99-101.
15. See N. Novacco, 'Problemi delle regioni sottosviluppate', and F. Compagna, 'Lo sviluppo economico del Mezzogiorno', in *Tempi Moderni*, October-December 1962, pp. 104-119 and 129-134.
16. Giuseppe Caronia and Roberto Calandra, 'L'attività urbanistica in Sicilia', in *Ordinamento regionale e pianificazione urbanistica*, Turin: Istituto Nazionale di Urbanistica, 1964, pp. 153-160.
17. See the report of the Martuscelli Commission published in *Urbanistica*, December 1966, pp. 29-160, i-li.
18. *Mondo Economico*, 19 May 1956, p. 4; 8 September 1956, pp. 7-9. See also Gianni Pirrone, 'Pianificazione economica e sociale', *Civiltà degli Scambi*, April 1958, pp. 46-49, and Francesco Renda, 'Il piano quinquennale per lo sviluppo economico e sociale della Sicilia', *Cronache Meridionali*, September 1956, pp. 536-543.
19. *Mondo Economico*, 8 May 1965, special issue on 'Sicilia 1965' and the second development plan.
20. *Mondo Economico*, 24 April 1965, special issue on the Sardinian plan; 'Il piano di rinascita della Sardegna', *Cronache Meridionali*, January-February 1960, pp. 47-57; Massimo Annesi, 'Origini, struttura e finalità, della legge sul piano di Rinascita della Sardegna', in *Le leggi sul piano di rinascita*, Rome: svimez, 1963, pp. 9-17; and Mario Berlinguer, 'La lotta per l'impostazione democratica del piano di rinascita della Sardegna', *Comune Democratico*, May 1961, pp. 136-138.
21. There seems to be no very satisfactory way to translate the Italian noun *urbanistica*, or, even worse, the adjective *urbanistico*. While the equivalents for the first include such expressions as 'city planning', 'town planning', 'urban development', 'physical planning', or the science or theory relating to those things, there seems to be lacking in English any adjective equivalent to *urbanistico*. For this reason, I use the neologisms 'urbanistics' and 'urbanistic' in order to convey in translation the full ambiguity of the original.
22. See the reports on the territorial plans presented to the 1952 congress of IMU in the volume entitled, *La pianificazione regionale*. The ministerial guidelines were published in two volumes entitled, *I piani territoriali: Criteri di indirizzo per lo studio dei piani territoriali di coordinamento* in Italian, 2 vols., Rome: Ministero dei Lavori Pubblici, 1952-1953.
23. Giacomo Corna-Pellegrini, 'Esperienze' (cited above, n. 1), pp. 435-425; Manin Carabba, 'L'articolazione territoriale della programmazione', *Tempi Moderni*, July-September 1963, 97-108 at p. 99; Umberto Toschi, 'Piani territoriali e piani economici', *Rassegna Economica*, October-December 1958, pp. 667-685; Ministero dei Lavori Pubblici, 'I Piani territoriali di coordinamento in Italia: Evoluzione de studi ed esperienze', in the proceedings of the IX National Congress of IMU, 1962, pp. 6-69.
24. Giorgio Mugnaini, 'Costituzione e precedenti dei comitati per la programmazione economica', *La Regione*, December 1964, pp. 281-89 at pp. 282, 288; E. Colombo, *Indirizzi* (cited above, n. 7), pp. 62-63, 137-139; M. Carabba, 'L'articolazione, delle regioni', *Cronache Meridionali*, December 1959, pp. 839-847; Giovanni Terranova, 'Gli schemi regionali dell'on. Colombo', *Nord e Sud*, February 1960, pp. 69-72.

25. E. Colombo, *Indirizzi* (cited above, n. 7), pp. 64-65.
26. See above, n. 23.
27. 'Gli incredibili strumenti della programmazione', *Questitalia*, November-December 1964, pp. 49-53; July 1965, pp. 33-35; November 1965, pp. 34-37.
28. G. Mugnaini, 'Costituzione' (cited above, n. 23), pp. 283-285.
29. Federico Orlando, 'Dal Piano Vanoni ai Piani Colombo', *Fortuna*, December 9-24, 1959, pp. 259-60; Alfredo Testi, 'Le Camere di Commercio e la Programmazione Regionale', *Nord e Sud*, March 1965, pp. 52-57; *Il Sole*, 5 December 1965, p. 1.
30. G. Mugnaini, 'Costituzione' (cited above, n. 23), pp. 283-84.
31. 'Gli incredibili strumenti' (cited above, n. 26), *passim*; *L'Unità*, 12 September 1965, p. 4; Umberto Dragone, 'Il progetto di riforma del Ministero del Bilancio e l'istituzione dei comitati regionali per la programmazione', *Il Mulino*, October 1965, pp. 905-913.
32. These are particularly well reported in the issues of the magazine *Esperienze Amministrative* (Milan).
33. Carlo Beltrame, 'Gli istituti regionali di ricerche socio-economiche', *Esperienze amministrative*, June 1964, pp. 66-109.
34. Ibid, pp. 69-70; the references in n. 11; Giuseppe Grosso (Pres. of the Prov. of Turin), 'Le funzioni delle Amministrazioni provinciali e delle Camere di Commercio', *Comune Democratico*, February-March 1962, pp. 45-46; Vitantonio Lozupone (Mayor of Bari), 'Formazione dei piani regionali', *Civiltà degli Scambi*, September-October 1959, pp. 46-47.
35. Beltrame, 'Gli istituti' (cited above, n. 32), pp. 74-75.
36. A. Testi, 'Le Camere di Commercio' (cited above, n. 28). See also the *Notiziario* of the Ufficio per la Programmazione e l'Informazione Economica of the Unione Italiana Camere di Commercio. Descriptions of the operations of some Regional Planning Committees can be found in Claudio Simonelli, 'L'esperienza del comitato regionale di programmazione economica del Piemonte', *Tempi Moderni*, Autumn 1967, pp. 141-152, and in Calogero Muscarà, 'I primi passi della programmazione regionale', *Nord e Sud*, August-September 1967, pp. 63-75.
37. There is also the delicate matter of financing the regional governments still to be settled. The Regional Electoral Law approved by the Italian Senate on 14 February 1968, requires the Government to submit to Parliament a bill on regional finance before the regional assemblies are elected. On the physical implications of the Five-Year Plan, see Franco Archibugi, 'L'assetto territoriale nella programmazione economica', *Urbanistica*, March 1967, pp. 4-7. The 'bridge-law' of August 1967 is treated in Giovanni Astengo, 'Primo Passo', *ibid*, October 1967, pp. 3-4, and Michele Martuscelli, 'La legge ponte: significato e operatività', *ibid*, pp. 5-13. The parliamentary debate on the regions is examined by Elio Rogati, 'Pro e contro le regioni', *Nord e Sud*, January 1968, pp. 38-50.
38. A preliminary 'treaty' between the Ministry of the Budget and the Ministry of Public Works, assigning roles and jurisdictions to the field agencies of each ministry, is contained in the joint circular of 23 July 1966, entitled, 'Elaborazione degli schemi regionali di sviluppo e dei piani territoriali di coordinamento'. The circular is reprinted in *Urbanistica*, March 1967, pp. 117-118. The entire March 1967 issue deals with the problem of co-ordinating economic and physical planning at the regional level.

# 33 Regional Development Policy in Spain

H. W. Richardson

## Introduction

A survey of regional policy in Spain is worth while for several reasons. In the first place, at a time when a considerable amount of work has been published on regional problems and policies in many countries of Europe and elsewhere there is still no work published in English on what has happened in Spain. Secondly, most of the countries studied are either liberal democracies or centrally planned economies; Spain is neither. She has adopted some degree of planning, indicative sectoral planning at the national level, and is preparing a Third Development Plan (1972-75). Despite the political complexion of the government, there is a strong trend in economic policy in favour of more emphasis on market forces, and this market orientation makes it very difficult to implement an active regional policy. Other factors are that regionalism is very strong in Spain and that pressure group activity, though not absent, takes a very different form compared with the rest of Western Europe. Thirdly, analysis of regional development problems is particularly interesting in countries at an intermediate level of development. Spain has a high aggregate growth rate, and the pressure for fast growth is very strong; at the same time, regional income differentials are very wide and the need for convergence is great. Thus, the con-

flict between efficiency (fast growth) and equity is very intense.<sup>1</sup> This conflict is resolved in Spain largely by subjugating the needs of regions to the pursuit of national development. Fourthly, Spanish regional policy is largely a growth pole policy. It reflects the influence that Perroux and other growth pole theorists in France have had on Spanish economists and also draws upon French and Italian planning experience (Comisaria del Plan, 1968-70).

The regional problem in Spain does not manifest itself as a classical north-south dualist problem. It is true that the three main industrial areas (Barcelona, Madrid and Cantabria, including Oviedo, Santander, Bilbao, San Sebastian and Vitoria) are in north and central Spain, and that even outside the industrial areas incomes per head tend to be higher in the north than in the south. But there are other concentrations of population and economic activity in other parts of the country, particularly around Valencia and Sevilla. More striking than the north-south dichotomy is the fact that the main centres of activity are so far apart from each other and are linked by an imperfectly developed transport network. This largely reflects topographical features. The most obvious physical characteristic of Spain is a large, high central plateau (meseta) or desert, encircled by rugged mountains often extending to the coastline. The result

<sup>1</sup> Of course, such conflict need not arise if the low-income regions offer higher rates of return, for in this case pumping more investment into poor regions will both raise efficiency and improve equity. Unfortunately, this condition does not hold in Spain.

is that development has taken the form of expansion in a number of widely separate regions, located (apart from Madrid) in the coastal lowlands between which there are areas of sparse population. The main population concentrations are at least 250 miles apart from each other. In economic terms this multiple regional disintegration reveals itself in the fact that most industry is oriented to regional rather than national markets and in wide inter-regional differences in incomes, prices and levels of consumer durable ownership. In these circumstances achieving a closer integration of the national economy is quite an important regional policy objective.

Regional policy in Spain is of fairly recent origin. The first legislation dates from January 1964 when the 'polos' were set up as part of the First Development Plan.<sup>2</sup> The strategy of regional development has always been seen as a dimension of a national growth policy, a geographical counterpart to sectoral planning. Regional equity objectives (such as the need to narrow interregional *per capita* income differentials or to reduce out-migration from rural regions) that figure so prominently in regional policies in other countries have in Spain crept in only surreptitiously as one of the goals of policy. The need to conform to efficiency criteria and to maximise the aggregate growth rate have remained the dominant official goals. This philosophy stems from the report of the I.B.R.D. Mission which visited Spain in 1962 and prepared the way for the First Plan (I.B.R.D., 1963). The Mission was very doubtful indeed about committing resources to an active regional policy, on the grounds that only a massive pumping of government expenditure would get results and this could not be afforded in Spain. Moreover, there were many very productive investment opportunities elsewhere in Spain that could boost the rate of growth in the national economy.

Thus, regional development policy ought to be subject to efficiency criteria and pursued only when it did not interfere much with the overall growth objectives. This would mean a careful selection of areas with sound development potential, and would involve choosing zones of development smaller in size than a province, likely to be situated in regions already having a fair degree of development. This suggestion clearly anticipated the 'polos' policy. In other regions with more limited prospects, the government should maintain tolerable living standards through social expenditures and by inducing out-migration. Secondly, having chosen the places for development, measures should be adopted that do not interfere with growth elsewhere. The Mission considered that this meant stimulating industries in those regions that were labour-intensive and which had no serious cost disadvantages relative to markets and suppliers (I.B.R.D., 1963, Ch. 15). This principle, that maximising the overall growth rate is the most satisfactory way of increasing employment and reducing inter-regional disparities and that departure from this goal is justifiable only in cases of great hardship or when the economic costs are very small has determined the strategy and character of regional policy in Spain. There have been shifts in emphasis giving a little more priority to regional equity objectives between the First (1964-67) and the Second Plan (1968-71), and possibly this shift will move a little further in the Third Plan. Nevertheless, the ranking and primacy of the efficiency/growth goal have remained unchanged.

The regional policies pursued refer to two different types of area. First, they deal with the intensive development of specific locations, such as poles of development ('polos de desarrollo'), centres of decongestion ('poligonos de descongestion') and zones of preferential location. This is the most important aspect of regional

<sup>2</sup> It is true that before 1964 there were regional programmes in existence such as the Badajoz and Jaen Plans. But these were not considered likely to attract much industry to these regions. Instead, they attempted to improve agriculture, e.g. through irrigation schemes, and to stimulate the development of processing industries and auxiliary services related to agriculture. Even so, the costs of this limited programme were high relative to the results achieved.

Table 1

| REGIONAL INDICATORS IN THE PROVINCES OF SPAIN |   |  |   |                        |  |         |
|---|---|--|---|------------------------|--|---------|
|   | Per capita<br>Income, 1967<br>('000 pesetas,<br>current prices) | Annual Rate of<br>Income Growth,<br>1964-67<br>(constant prices) | Migration,<br>1960-67<br>as % of<br>1960 Population | Activity<br>Rate, 1967 | Agricultural<br>Total<br>Employment<br>Ratio, 1967 | Region* |
| Vizcaya                                       | 68.1  | 3.67   | +13.4   | 43.4                   | 9.5  | 3       |
| Madrid  | 67.9  | 1.96   | +7.5  | 39.7                   | 4.1  | 5       |
| Guipúzcoa                                     | 66.5  | 3.79   | +7.7  | 45.4                   | 7.5  | 3       |
| Barcelona                                     | 63.3  | 4.07   | +17.7   | 46.0                   | 3.8  | 8       |
| Gerona  | 59.9  | 4.30   | +5.8  | 52.2                   | 19.1   | 8       |
| Baleares                                      | 58.7  | 4.49   | +1.7  | 50.4                   | 22.8   | 8       |
| Navarra                                       | 55.9  | 3.69   | +2.2  | 46.3                   | 29.3   | 7       |
| Alava   | 55.2  | 2.59   | +12.9   | 45.4                   | 19.6   | 3       |
| Logrono                                       | 52.9  | 4.70   | -1.9  | 46.3                   | 40.8   | 7       |
| Santander                                     | 51.3  | 6.11   | -0.8  | 46.4                   | 32.0   | 2       |
| Lerida  | 50.1  | 5.39   | -1.6  | 47.1                   | 42.0   | 8       |
| Tarragona                                     | 49.4  | 3.47   | +5.5  | 46.5                   | 35.2   | 8       |
| Zaragoza                                      | 47.1  | 5.57   | +2.2  | 42.0                   | 29.0   | 7       |
| Huesca  | 46.7  | 6.46   | -4.3  | 44.3                   | 40.8   | 7       |
| Burgos  | 46.6  | 6.74   | -7.1  | 48.3                   | 46.1   | 4       |
| Valencia                                      | 46.1  | 4.85   | +7.1  | 41.6                   | 32.6   | 9       |
| Palencia                                      | 45.9  | 7.65   | -9.0  | 44.1                   | 42.4   | 4       |
| Valladolid                                    | 45.8  | 5.14   | -0.8  | 39.7                   | 29.7   | 4       |
| Castellon                                     | 44.4  | 4.36   | +4.2  | 48.7                   | 40.1   | 9       |
| Oviedo  | 42.5  | 6.74   | -0.4  | 41.1                   | 33.6   | 2       |
| Segovia                                       | 41.2  | 7.74   | -8.3  | 40.0                   | 47.6   | 4       |
| Soria   | 40.8  | 8.46   | -12.7   | 40.6                   | 51.4   | 4       |
| Teruel  | 40.8  | 7.77   | -14.5   | 48.5                   | 57.0   | 7       |
| Leon  | 40.1  | 6.86   | -3.8  | 46.3                   | 47.6   | 4       |
| Alicante                                      | 38.0  | 4.05   | +5.1  | 36.2                   | 21.0   | 9       |
| Guadalajara                                   | 37.1  | 8.28   | -10.2   | 44.1                   | 52.9   | 6       |
| Pontevedra                                    | 35.2  | 3.92   | -0.6  | 44.8                   | 45.2   | 1       |
| Zamora  | 34.9  | 7.92   | -8.8  | 45.0                   | 61.9   | 4       |
| Salamanca                                     | 34.8  | 7.27   | -7.2  | 38.8                   | 50.4   | 4       |
| Albacete                                      | 33.4  | 7.04   | -12.9   | 37.3                   | 49.5   | 6       |
| Toledo  | 32.9  | 7.65   | -7.3  | 37.4                   | 52.6   | 6       |
| Cuenca  | 32.8  | 8.18   | -15.5   | 37.8                   | 61.8   | 6       |
| Murcia  | 32.7  | 5.70   | -2.6  | 32.6                   | 37.4   | 9       |
| Sevilla                                       | 32.5  | 4.98   | -4.4  | 32.4                   | 35.7   | 10      |
| Malaga  | 31.9  | 5.21   | -4.3  | 33.6                   | 35.6   | 10      |
| Cadiz   | 31.8  | 5.45   | -3.3  | 30.8                   | 25.7   | 10      |
| La Coruna                                     | 31.2  | 5.67   | -0.9  | 41.0                   | 51.1   | 1       |
| Las Palmas                                    | 31.2  | 2.43   | +0.7  | 30.5                   | 38.0   | 11      |
| Ciudad Real                                   | 31.2  | 7.64   | -10.5   | 36.2                   | 46.1   | 6       |
| Huelva  | 31.1  | 6.16   | -5.5  | 33.4                   | 35.3   | 10      |
| Lugo  | 31.0  | 6.68   | -4.6  | 46.8                   | 66.1   | 1       |
| Avila   | 30.0  | 6.76   | -6.3  | 42.1                   | 66.2   | 4       |
| Santa Cruz de Tenerife                        | 29.2  | 2.75   | +0.4  | 31.1                   | 44.0   | 11      |
| Cordoba                                       | 28.8  | 7.24   | -11.0   | 34.5                   | 50.2   | 10      |
| Orense  | 27.3  | 6.26   | -2.5  | 45.2                   | 66.0   | 1       |
| Granada                                       | 26.8  | 6.84   | -10.3   | 34.5                   | 53.7   | 10      |
| Caceres                                       | 25.5  | 7.43   | -11.3   | 35.7                   | 57.5   | 6       |
| Badajoz                                       | 25.4  | 7.36   | -11.2   | 33.1                   | 55.6   | 6       |
| Jaén  | 24.4  | 7.07   | -11.2   | 34.9                   | 54.4   | 10      |
| Almeria                                       | 23.1  | 5.00   | -8.2  | 31.8                   | 49.1   | 10      |
| ESPANA  | 44.7  | 5.31   | ±3.6  | 40.2                   | 31.1   |         |

\* The regional classification is as follows: 1—Galicia; 2—Cantabrica; 3—Vasca; 4—Castellano-Leonesa; 5—Madrid; 6—Extremeno-Manchega; 7—Aragonesa-Riojana; 8—Catalana-Balear; 9—Levantina; 10—Andaluza; 11—Canarias.

policy, and the one where efficiency, national integration and regional equity goals intermesh. Secondly, in some cases the policymakers are concerned with the global treatment of a region. These are the 'regiones en crisis', agricultural regions with below average rates of growth such as the coast of Gibraltar, Tierra de Campos, Badajoz and Jaén. The method of dealing with these is the preparation of a specific plan including agricultural development and public investment schemes. As far as attracting industry is concerned, these regions are unlikely to have much success in the foreseeable future since the level of inducement offered is much the same as at the poles which are much more attractive locations for development. Intervention in the rate of progress of these regions is justifiable solely in terms of equity objectives, since they are in the main isolated and their limited market potential throws up few investment opportunities.

#### Regional Income Differences

Despite its impressive growth performance in recent years (6% average annual rate of growth in G.N.P.), Spain retains many of the characteristics of a non-industrial economy. The average *per capita* income level remains low (less than £270 per annum in 1967). Urbanisation has not proceeded very far for a country of 34 million; only four cities have a population greater than 500,000, four more are in the size class 250,000-500,000, while 38% of the population live in places of below 10,000. Most provinces are still predominantly agricultural, with 36 out of 50 having more than one-third of the labour force employed in agriculture while in 16 provinces the agricultural share is over 50%.

It is characteristic of countries at an intermediate stage of development that they experience (1) wide inter-regional income differentials and (2) high rates of rural-urban migration. Spain is no exception. The basic available data

are assembled in Table 1. They are presented for the 50 Spanish provinces; although column 6 shows how these may be aggregated into 11 regions, the regional breakdown is somewhat arbitrary since there is no regional organisation as such and classification of regions according to homogeneity or polarisation criteria conflicts with traditional cultural divisions. The poorest region has a *per capita* income level only 34% of that of the richest. In the period 1960-67, 6 provinces had a positive and 18 provinces a negative migration rate in excess of 7%. In more general terms, there are, as shown in Table 2, strong correlations between *per capita* income levels and migration rates (positive), activity rates (positive) and labour shares in agriculture (negative). Much of the explanation of regional income disparities can be decomposed into two factors: wide inter-regional differentials in agricultural productivity; and significant regional differences in economic structure, especially the relative importance of manufacturing (and to a lesser extent service) employment.<sup>3</sup>

Table 2

CORRELATION COEFFICIENTS: REGIONAL INDICATORS  
FOR 50 SPANISH PROVINCES

|   |          |          |          |         |
|---|----------|----------|----------|---------|
| 2 | -0.5056* | ...      | ...      | ...     |
| 3 | 0.7158*  | -0.8301* | ...      | ...     |
| 4 | 0.6537*  | -0.0703  | 0.3424   | ...     |
| 5 | -0.7836* | 0.7191*  | -0.8178* | -0.1519 |
|   | 1        | 2        | 3        | 4       |

\* Significant at the 0.001 level.

1—*Per capita* income. 2—Rate of growth in *per capita* income. 3—Migration rate. 4—Activity rate. 5—Agricultural/total employment ratio.

The wide spread of regional income differentials should not be taken to imply increasing disparity. In fact, there have been strong signs of regional convergence in recent years. Some indication of this is given by the annual rates of growth in *per capita* incomes shown in column

<sup>3</sup> The regression equation relating *per capita* income to the agricultural labour share for the provinces takes the form:

$$\frac{Y}{P} = 65.04 - 0.602 \frac{A}{E} \quad (R^2 = 0.6141).$$

(3.00) (0.069)



(2) of Table 1. The nine wealthiest provinces all had rates of growth below the national average, while of the 32 provinces with a *per capita* income below the national median level 24 had above average rates of growth (and 10 of these had a growth rate more than double that enjoyed by the wealthiest province—Vizcaya with a growth rate of 3.67%). It is true that convergence forces were somewhat stronger in the period 1964-67 than in the previous years, largely because the agricultural sector performed somewhat better during this period than the industrial sector which experienced two years of recession. The Gini coefficient for regional *per capita* income distribution had in fact increased somewhat in the period 1955-64 from 0.249 to 0.2528 (though it fell again to 0.2369 in 1967). Nevertheless, even in the 1955-64 period the rich provinces did not grow fastest; instead, the fastest growing provinces were largely those in the middle range.<sup>4</sup>

Out-migration from the low-income provinces was the major factor making for regional *per capita* income convergence.<sup>5</sup> The growth rates in *total* output do not vary between provinces nearly so much as those in *per capita* income, and they strongly favour the high-income regions; output increased faster in Vizcaya, Guipúzcoa, Madrid and Barcelona than anywhere else (Presidencia del Gobierno, II Plan, 1969, p. 117). The equalising tendencies were obtained only by rapid shifts of population which swelled the populations of the metropolitan and industrial areas and helped to relieve overpopulation in the rural areas. Thus between 1960 and 1967, when the population of Spain increased by 6.7%, the wealthy regions experienced much faster population growth: the proportionate increase was 31.7% for Alava, 30.5%

for Madrid, 28.7% for Vizcaya, 25.1% for Guipúzcoa, and 23.1% for Barcelona. Yet population declined *absolutely* in 23 out of 50 provinces, and the fall exceeded 10% in 12 of these. The absolute changes were also very large: population fell in the seven provinces with the most marked absolute depopulation by 465,000 but rose in the seven provinces with the heaviest population absorption by 2,238,000 (an amount greater than the *total* population increase of Spain—2,025,000). On the face of it, this gravitation of population suggests that the link between migration and economic opportunity was the main force making for regional *per capita* income convergence. One fact which rather spoils this neat hypothesis is that activity rates in all the rich regions with the exception of Barcelona fell between 1964 and 1967. This is possibly explained by recession and a tendency for in-migrants to overestimate employment opportunities.

The evidence for capital movements is very scanty indeed, but the indicators available suggest that capital flows did not have an equilibrating role and that the brunt of convergence must have fallen on labour. In the first place, government investment did not favour the south and the poorer provinces. Secondly, as far as private capital movements are concerned, industrial investment was strongly concentrated in the established industrial areas. Some observers, such as Lasuen (1962), have argued that the banking system, with its involvement in industry, has operated in such a way that deposits in the poor regions are transformed into credits in the north. Similarly, the wealthier groups with high propensities to save living in the south tend to invest in the developed regions rather than nearer home.

<sup>4</sup> They include Huesca, Gerona, Lerida, Baleares, Soria, Toledo, Granada, Albacete, Malaga and Alicante. The gap between the richest and poorest region narrowed considerably between 1955 and 1967. In 1955 *per capita* income in the richest region was more than four times that in the poorest; by 1967 the ratio had fallen to less than 3 : 1.

<sup>5</sup> The regression equation relating the rate of growth in *per capita* income to the migration rate in the 50 provinces takes the form:

$$\frac{\Delta Y/P}{Y/P} = \frac{5.214 - 0.189}{(0.146) (0.018)} \frac{M}{P} \quad (R^2 = 0.6891).$$

One other factor needs a brief mention. It is quite widely believed that the tourist industry in Spain is crucial to the prosperity of many regions. This statement requires considerable qualification. In only ten provinces does tourism account for more than 4.5% of gross value added, and in only two for more than 10%: Baleares (20%) and Gerona (13.7%). Apart from these two provinces which do enjoy high *per capita* incomes, the next four most tourist-minded (Malaga, Granada, Las Palmas and Santa Cruz de Tenerife) are in the bottom third of the *per capita* income league table. Furthermore, Spanish economists frequently complain that the net income flow from abroad due to tourism is largely illusory because so much of the income created (apart from wage payments in the service industries) flows abroad again. This is because a high proportion of tourist infrastructure is in foreign hands.

#### Policy Instruments

There are three main instruments of regional policy in Spain:

- (1) an investment grant ('subvención') of 20% in the poles of industrial promotion and in preferential zones of industrial location in the rural regions and of 10% in the poles of development;
- (2) certain fiscal exemptions (reductions up to 95% in taxes and import duties on machinery and equipment not manufactured in Spain; free depreciation allowances during the first five years);
- (3) preference in obtaining official credit.

Of these, the latter is easily the most important, accounting for over two-thirds of government expenditure on financial incentives to firms. In addition to these three direct inducements to firms, there is a certain amount of infrastructure investment at locations selected for development which may have some attractive power. However, this infrastructure is financed out of the budget allocations of several ministries, and is not explicitly linked with regional policy measures.

Although some of the regional policy instruments employed in Spain are unusual, it would be wrong to criticise measures that are part of the country's institutional environment and that have frequently been borrowed from other fields of economic policy simply because they are unfamiliar. However, the *effects* of these measures can still be subject to scrutiny, particularly on whether they conform to the goals of policy and to the needs of Spanish regional development.

The first thing striking the observer is that the range of policy instruments is a broad one, and the mix of investment subsidies, easy credit and tax exemptions makes it difficult for firms to calculate the precise value of the benefits they receive. Secondly, given the scarcity of budgetary resources in Spain and the dominance of the efficiency goal in regional policy, it is important that measures should be cost-effective in terms of both fiscal cost and resource cost. Up to the present, there has been no attempt in Spain to subject their regional policy instruments to cost-effectiveness criteria. Thirdly, the Spanish authorities have not yet resorted to negative controls on industrial development in the major metropolitan centres as a means of steering industry to other regions. Fourthly, all the policy instruments in operation have a capital-intensive bias. Although this might be defended on the grounds that they promote a high rate of investment which is important for growth in Spain and that stimulating capital-intensive techniques at the poles may induce the spread of more advanced technology and higher productivity, it is quite inappropriate for an economy suffering from capital shortage and an excess supply of labour. It would be more sensible to have policy measures that promote the demand for labour. Assuming at least some flexibility in the choice of factor utilisation ratios, subsidies with a capital-intensive bias must induce firms to adopt a higher capital-labour ratio than is efficient given the relative factor price ratios.

Some comment is also necessary on individual measures. The provision of official credit to mobile firms absorbs a high proportion of total

expenditure on regional development, and is regarded as essential because of the low degree of self-financing in Spanish industry and because the shortage of liquidity is too frequently a severe constraint on the expansion of firms. Yet the flow of official credit, and hence the effectiveness of regional policy, fluctuates a great deal because of the failure to insulate the Bank of Industrial Credit from monetary movements in the economy as a whole and from the impact of cyclical fluctuations. Moreover, small firms have been relatively deprived of aid from this source.

As for tax exemptions and import duty relief, these are probably unsound as regional policy instruments because they may distort the input mix of the production technology of firms in receipt of these benefits. In particular, firms adopting a short time horizon may choose a technological structure that is the least-cost method for the limited period of the tax relief but becomes inefficient when they have to pay the full cost of their inputs.

There are disadvantages to the policy of investment subsidies ('subvenciones') as applied in Spain. First, it is at least possible that a 10% investment subsidy is too low to achieve the momentum of inter-regional mobility required. Secondly, the subsidies are offered only at annual 'concursos' and are not available at all times to intending applicant firms. This inevitably means a lack of continuity and flexibility. Moreover, there are doubts about the criteria on which discriminatory decisions have been made as to which firms should receive aid. Two points in particular deserve comment. Only industries on a select list drawn up for each pole or area can receive a subsidy, but in the absence of a specific use of industrial complex analysis the range of industries is so broad that inequities arise in respect of activities excluded from the list. It is probably harmful to deny assistance to any firm *merely* on the grounds of its industrial classification. Recently, there has been a change in eligibility rules which denies small investment projects the chance of a subsidy. Although it is true that not enough large projects have been

attracted to the poles, this is the wrong way to solve the problem. There are three main reasons for this view: even if large firms are the main generators of development at the poles small and medium-sized firms will continue to be needed as auxiliary and supply plants; small firms tend to be more labour-intensive and the capital cost per job low, a tendency that could help to correct the capital-intensive bias of existing measures; this restriction will undoubtedly discriminate against the growth of Spanish indigenous industry which is largely small-scale in structure.

### The 'Polos de Desarrollo' Policy

In so far as it is concerned with industrial development, Spanish regional policy is predominantly a poles of development policy. Early in 1964, seven poles were established—Zaragoza, Sevilla, Valladolid, La Coruña, Vigo, Burgos and Huelva. The first five were designated 'polos de desarrollo' while the latter two were (up to the end of 1968) 'polos de promoción industrial'. Firms were offered the same inducements to go to both, except that the initial investment grant for the industrial promotion pole was 20% rather than 10%. The poles hold their status for only a limited period (Zaragoza 6 years, Sevilla and Valladolid 7, La Coruña and Vigo 8, and Burgos and Huelva 5 as industrial promotion poles followed by 5 as poles of development). In 1969 another four poles were designated (Granada, Cordoba, Oviedo and Logroño) initially for five years, and these will come into operation in the years 1970-72 (Fig. 1).

A pole strategy is probably suited to Spanish conditions. In a country with limited investment resources for regional development it makes good sense to concentrate these resources at a limited number of centres, either at points with natural locational advantages or where it is feasible to generate a sufficient build-up of investment to obtain agglomeration economies. Furthermore, in a country with backward regions in which there are socio-economic obstacles to development a growth pole approach is one method of



Figure 1

attempting to diffuse 'growthmindedness' into these areas. Finally, since promotion of national integration is a goal of regional policy, this probably requires development axes radiating out of Madrid to be strengthened by the location of thriving centres along the axes. A pole strategy could be a great help here.

The most striking fact about the poles is that they are not concentrated in the most backward regions. The original poles were located in the provinces ranked 13, 15, 18, 27, 34, 37 and 40 in *per capita* income terms, while the more recent designations are in provinces 9, 20, 44 and 46. Although the creation of poles at Cordoba and Granada suggests a shift in favour of equity

objectives, it still seems clear that the poles are regarded as an extension of national and sectoral planning strategy (as in French planning practice) more than as an instrument for developing the poorest regions. An important criterion of pole selection has been to find suitable centres for expansion along major (present or intended) transport and development axes. The selection of Zaragoza, Valladolid and Sevilla fall into this pattern, and more recently Cordoba cements the Madrid-Sevilla axis, Oviedo helps to link Santander-Bilbao with La Coruña and Logroño lies along a major axis from Madrid to Guipúzcoa. On the other hand, several of the original poles (La Coruña-Vigo, Valladolid-Burgos,

Sevilla-Huelva) were too close together to be justified in terms of a grand strategy for a national hierarchical network of urban centres. In most cases, regional equity considerations were not very important. Had they been so, the obvious procedure would have been to choose particular problem regions first, and subsequently to select the best places for development within these regions as a pole. The choice of Granada probably reflects the equity goal most, since its comparative locational advantages are probably non-existent and it is not located on a major development axis. Yet it is doubtful whether attracting industry there is a more appropriate solution than a modest rate of out-migration, since in the 1960s out-migration from the province was associated with a rise in income per head faster than the national average.

As yet, there is insufficient evidence available to permit a proper evaluation of the success or failure of the poles strategy. However, some of the basic statistics on investment, employment, subsidies and industrial distribution for the first five years of the policy have been assembled in Table 3. The information in the table largely speaks for itself, but it may be useful to highlight a few points. In relation to their size, Burgos and Huelva obtained high shares of the total investment. Although this may reflect special locational advantages, access to raw materials in the case of Burgos and a good location and communications relative to external markets for Huelva, there is *prima facie* evidence that the higher investment subsidy offered to the poles of industrial promotion made a substantial difference. Investment expenditure per job was exceptionally high in Huelva compared with the other poles, but this was primarily due to the high capital intensity of the chemical and energy industries rather than to any inherent locational disadvantages. Vigo was much more successful than its near neighbour La Coruña in attracting industry, particularly because of a thriving metal products sector while almost a third of the total investment in La Coruña went into the energy sector which was so highly capital-intensive that

only 72 jobs were created at an investment cost of 15.2 million pesetas per job. Of the poles, probably Valladolid would pass efficiency tests most easily. Average size per installation in terms of employment was high (one of the failures of the poles policy as a whole was that it attracted too many small-scale enterprises) yet investment per man was below average, but most important of all the budgetary cost of inducing expansion at Valladolid was very low (only 15.2% of total investment). It is, of course, the nearest pole to Madrid, though not too close to suffer from polarisation effects.

Apart from these specific points, there are some more general observations. First, the number of jobs created by the poles policy, 57½ thousand, is really quite small, equivalent to 3.1% of the total population of the poles, or, putting it another way, 0.4% of the labour force of the country as a whole. Moreover, many of these jobs cannot be regarded as *net* additions to the labour force induced by policy but were probably diverted from elsewhere. Secondly, the availability of official credit was, with the exception of Huelva, several times more important than investment subsidies (the average ratio was greater than 3 to 1). Thirdly, the main industries attracted to the poles were concentrated in a few sectors; chemicals and metal products accounted for one-half of the investment and construction and food industries for a further quarter. Finally, there are a number of important facts not brought out in the table. Some poles had a marked 'backwash' effect on nearby areas, particularly Sevilla on the rest of Andalusia, though Huelva had a similar effect as did Valladolid in the north. Most of the entrepreneurial initiative at the poles was non-local and came from advanced regions such as Madrid, Cataluna and Vasco, though foreign capital played a fairly small role in pole development. A high proportion of the firms attracted to the poles were small scale; as many as 56% required an investment of less than 50 million pesetas, and only 22% over 100 million.

Some aspects of the poles policy give rise to inconsistencies and difficulties. For instance,

Table 3  
STATISTICS OF POLE DEVELOPMENT, 1964 TO JUNE 1969

|            | 1                    | 2                       | 3           | 4                                      | 5                         | 6                                    | 7                               | 8                                     | 9   | 10                      | Industrial Distribution of Investment (%) |          |               |       |           |              |              |                |        |           |               |
|------------|----------------------|-------------------------|-------------|--|---------------------------|--------------------------------------|---------------------------------|---------------------------------------|---|-------------------------|---|----------|---------------|-------|-----------|--------------|--------------|----------------|--------|-----------|---------------|
|            |                      |                         |             |  |                           |                                      |                                 |                                       |   |                         | Food                                      | Textiles | Timber Trades | Paper | Chemicals | Construction | Basic Metals | Metal Products | Energy | Education | Other Sectors |
|            | No. of Installations | Investment (b. pesetas) | Jobs ('000) | Avg. Investment per Plant (m. pesetas) | Avg. Employment per Plant | Avg. Investment per Job (m. pesetas) | Subsidies to Firms (b. pesetas) | Official Credit to Firms (b. pesetas) | Government Aid as % of Investment (2÷[7+8]) | Population, 1967 ('000) |   |          |               |       |           |              |              |                |        |           |               |
| Burgos     | 85                   | 6.24                    | 9.99        | 74                                     | 117                       | 0.62                                 | 0.515                           | 1.526                                 | 32.8  | 110.0                   | 17.5                                      | 6.7      | 5.6           | 5.9   | 27.1      | 9.3          | -            | 25.7           | -      | 2.2       | -             |
| Huelva     | 46                   | 8.86                    | 4.02        | 192                                    | 87                        | 2.20                                 | 1.133                           | 1.617                                 | 31.0  | 98.0                    | 3.2                                       | 0.4      | 4.1           | 1.6   | 68.2      | 1.2          | 0.5          | 3.1            | 15.6   | 0.8       | 1.3           |
| La Coruña  | 36                   | 3.51                    | 2.65        | 98                                     | 74                        | 1.32                                 | 0.066                           | 0.532                                 | 17.1  | 211.6                   | 27.0                                      | -        | 1.3           | 4.6   | -         | 8.7          | 14.0         | 5.9            | 31.2   | 7.3       | -             |
| Sevilla    | 83                   | 8.33                    | 10.86       | 100                                    | 131                       | 0.77                                 | 0.433                           | 1.631                                 | 24.7  | 598.3                   | 10.0                                      | 15.9     | 1.0           | 5.3   | 11.9      | 22.1         | 2.1          | 18.7           | 7.6    | 5.4       | -             |
| Valladolid | 55                   | 8.30                    | 11.26       | 150                                    | 205                       | 0.73                                 | 0.201                           | 1.058                                 | 15.2  | 203.0                   | 14.9                                      | 0.8      | 0.6           | 7.8   | 2.3       | 9.8          | -            | 61.5           | 0.5    | 1.8       | -             |
| Vigo       | 57                   | 5.32                    | 9.67        | 94                                     | 170                       | 0.53                                 | 0.170                           | 1.141                                 | 24.7  | 196.1                   | 10.4                                      | 2.7      | 0.3           | 4.3   | 3.2       | 26.9         | -            | 46.1           | 3.3    | 2.8       | -             |
| Zaragoza   | 112                  | 6.08                    | 8.89        | 54                                     | 78                        | 0.70                                 | 0.270                           | 1.173                                 | 23.7  | 465.5                   | 10.4                                      | 0.5      | 0.3           | 11.3  | 5.0       | 14.3         | 2.9          | 40.9           | 5.2    | 8.9       | 0.3           |
| All Poles  | 474                  | 46.64                   | 57.34       | 99                                     | 121                       | 0.81                                 | 2.787                           | 8.679                                 | 24.6  | 1,822.4                 | 12.0                                      | 4.3      | 2.0           | 5.7   | 20.1      | 12.7         | 1.9          | 29.3           | 7.9    | 3.8       | 0.3           |

the degree of intervention demanded for a successful poles policy may be at odds with the prevailing ethos about the extent of government interference in overall economic policy. In Spain the government is trying to implement a growth poles policy at a time when the tide is moving in favour of more emphasis on market forces in the economy as a whole. If poles are regarded as *planned* centres (albeit with strategic advantages for national and/or regional development) it probably needs extensive intervention to co-ordinate infrastructure provision, ensure synchronisation of public and private investment decisions and possibly to discriminate in favour of certain firms considered ideal for the pole. A related point is that a poles policy requires a great deal of co-ordination between government departments, and this co-ordination is sorely lacking in Spain.<sup>6</sup> Growth poles involve interaction between urban and regional problems, problems that are handled by separate ministries. More important is the fact that a poles policy needs co-ordination of investment decisions in several sectors: industrial investment in the private sector which determines the rate of job creation; the provision of housing, road-building and public utilities; the infrastructure for education and other social services. It is clearly undesirable to let investment decisions in one sector get too far out of phase with the others, yet this risk is unavoidable without more co-operation among the ministries concerned. At present, the physical planning aspects of regional development receive too little attention compared with the industrial aspects. Problems of environmental planning, moreover, have not been considered at all,<sup>7</sup> and when in the future they are, as they must be, they will demand more co-ordination by more ministries than any other aspect of regional policy. Another reason why co-ordination is essential relates to the instruments of regional policy. There are several

alternative combinations of spending a given budgetary sum on attracting industry to the poles: subsidies to firms, provision of cheap credit, expenditure on site development, factory building and roads, tax relief, investment in social infrastructure. Some combinations may be more effective in one situation than in another, and there may consequently be advantages in maintaining maximum flexibility in how available government expenditure is to be deployed. With the present structure of government in Spain, this flexibility is inhibited by the fact that a switch from one kind of subsidy to another would involve transfers of budget allocations between government departments.

As we have seen, the benefit period for the poles is short. This conflicts with the generally held conception of poles as centres that need to be nurtured over a long period, while the continuation of such a policy might have unfavourable implications for the spatial development of the Spanish economy. The brief subsidy period is probably rationalised in terms of budgetary constraints and the infant industry argument. Whereas the former is easily understood the infant industry analogy should not be pushed too far in a regional context. The principle is that after five to seven years of aid the subsidised firm located at the pole should be able to compete on equal footing with firms located elsewhere. But a pole is much more than a firm or industry needing a little initial help. It is a centre of expanding population and of highly interrelated economic activities capable of self-sustained growth, and it is doubtful whether the momentum for this can be created in five or six years. The gestation period for a fully developed urban infrastructure is very long, while the smaller auxiliary firms that may need to be brought in to the pole may not be attracted within the benefit period and are much less likely to come once the inducements have lapsed.

<sup>6</sup> A striking instance of this lack of co-ordination is that several government departments each have their own separate and independent programmes for the development of industrial estates. This makes for duplication and the creation of estates at inefficient locations.

<sup>7</sup> The signs of this are clearly visible in some of the suburbs of Madrid and in many Spanish towns affected by industrial development.

Even more serious, what will be the long-run consequences of a short benefit period strategy? What happens when the designation period for the four new poles expires? Will a new set of poles be created for another five or six years, and then another set, and so on? In this case, the long-run effect would be that a large number of centres would have experienced modest expansions, and that either Spain would have a system with too many small centres or alternatively would have a hierarchical distribution similar (relatively if not absolutely) to that which exists at present. It is unlikely that this would be the most rational way to amplify the present urban hierarchy. Indeed, it may be much more efficient and much more conducive to faster economic growth to develop more intensively a limited number of poles over a much longer period.

The brief designation period for poles is symptomatic of a much broader preoccupation with the short run in regional planning in Spain. It is widely recognised that regional policy is primarily a long-term strategy, and that the objective of achieving marked reductions in regional *per capita* income differentials can be achieved only over decades. Yet regional planners in Spain tie themselves to the projection period associated with national planning (three to four years), probably because regional planning is treated as a dimension akin to sectoral planning within the national planning framework. But this restriction on their forecasting horizon can lead policymakers to incorrect policy decisions. The long run is not merely the aggregation of a series of short-run steps, and the optimal path of adjustment will vary according to the time horizon selected. It is clearly desirable that regional policymakers in Spain should look at their problems with a much longer perspective. An obvious first step would be to outline some of the regional implications of the growth of the Spanish economy into the 1980s and up to the year 2000. Although sketchy and speculative, exploration of some of the possible spatial projections would make it easier to make wiser investment decisions in the nearer future. There are several

alternative strategies of development (gradual spreading out from the north-east quadrant of Spain, development of a sufficiently large agglomeration to act as a counterweight in either the south-east or the south-west, linking of the coastal area with a subsequent spread of development into the hinterlands, development along major transport axes out of Madrid, and others), selection of a particular strategy may be dependent upon the level of development in the aggregate economy achieved by the terminal projection year, and different strategies will have different implications for the selection of poles and their benefit periods (as well as for other aspects of regional policy).

A recent change in poles policy is that subsidies are now offered over an area more extensive than the pole itself. This is a welcome change if, as should be the case, one objective of the poles is to stimulate expansion over the growth area as a whole. However, since in a fully developed pole there is likely to be a difference in the average size of firms at the pole and in the growth area, there is an element of contradiction between this reform and the other recent policy change allowing only relatively large-scale investment projects to benefit from subsidies.

Two aspects of public policy that should be integrated into a poles strategy are education and housing policy. Neither is considered to be relevant to regional development policy in Spain. If the aims of the growth pole strategy include the transformation of social attitudes to growth, the transmission of innovations from rich regions and the training of skilled manpower and managers, the provision of institutes of technical (and sometimes university) education is relevant to a poles policy. Housing is even more important. To the extent that the poles policy depends on attracting new labour to the poles by immigration either from the surrounding countryside or from other regions, its smooth operation depends on the provision of sufficient housing close to the jobs at prices (or rents) that the workers can afford. In Spain there is a critical housing problem, there is virtually universal



reliance on the private sector, and subsidies (where they exist) are offered to builders rather than to households. Subsidies to builders do not give enough control over the types (and prices) of housing built and do not ensure that houses are provided precisely where they are needed. Solutions to the housing problem probably all require a higher level of public spending. One improvement would be to give housing subsidies (or very cheap and low-deposit mortgage credit) to workers moving to and working in the pole. Another solution might be for the State to extend the scale of its direct housing provision, possibly via the establishment of a public or semi-public organisation, subsidised by the State, with the express aim of building houses at the poles.

An effective growth pole strategy in Spain undoubtedly needs a greater injection of foreign (i.e. overseas) capital. The size of firms in Spanish industry is in general too small to have the generating effects needed at the poles. A poles policy is much easier to implement if very large corporations with an abundance of managerial talent, access to the highest levels of technology and able to withstand early losses can be persuaded to set up branch plants of some size at the pole (Lasuen, 1969). Spanish industry is not able to do this. There are large firms only in the banking industry and the public services sector (e.g. electricity, the telephones industry). There is not one 100% Spanish-owned firm among the 200 largest manufacturing companies and over 80% of Spanish industrial firms employ less than five workers (Common Market Report, 1969). If large firms have a strategic function in pole development, more foreign-owned plants must be attracted to the poles.

Up to the present, foreign capital has had only a limited role in the development of the poles, though its share in industrial investment at the poles (14.1%) has been much higher than its share in total Spanish investment. The latter is very low, only 4% of the total, despite the emotive objections to external capital in some quarters in Spain. Moreover, even a doubling of foreign capital's share in total investment

would not impose a debt burden on the Spanish economy. Of course, there are obstacles in the way of boosting foreign capital's participation in the poles strategy, namely the retarded development of Spain's new industries and the incomplete development of a national market, though these obstacles are likely to recede over time. If more foreign capital could be attracted to the poles, this would have another advantage given the dominance of efficiency and growth objectives in Spanish regional policy. This is because, provided the foreign capital attracted is additive not substitutive, i.e. not diverted from more profitable use elsewhere, it would enable regional development to be pursued without conflicting with national growth objectives.

#### The 'Poligonos de Descongestión'

Another facet of Spanish regional policy, related to the poles of development strategy, is the 'poligonos de descongestión'. In 1964 5 poligonos were established ostensibly to relieve pressure from Madrid, by persuading industry to move out from the metropolis and perhaps ultimately by dealing with overspill population in this way. These 'poligonos' are Toledo, Guadalajara, Arando de Duero, Alcazar de San Juan and Manzanares. Decentralising firms willing to go in the 'poligonos' can obtain official credit and benefit from tax exemptions.

It is doubtful whether a decongestion policy is justified in Spain in present circumstances. Without detailed studies of social agglomeration costs, and none have been carried out in Spain, these costs can be exaggerated. The policy may have developed from applying generalisations derived from metropolitan experience abroad to Madrid and from excessive reliance on the single indicator of road traffic build-up. Even this is not serious by international standards, and there are, of course, ways other than decentralisation to control or meet the expansion of urban traffic. Moreover, Madrid is relatively free from the disadvantages for private industry found in many metropolitan cities in advanced countries.

Land is fairly plentiful and relatively cheap in the suburbs, there is an excess supply of labour and there is little congestion outside the central area. It would probably be impossible in present circumstances to persuade many people to move from Madrid to the 'poligonos' where the level of amenities is much lower than in Madrid, wage levels are lower, employment opportunities for non-heads of households while scarce in Madrid are non-existent, and housing remains expensive.

In addition, at least three of the 'poligonos' (Arando de Duero, Alcazar de San Juan and Manzanares) are much too far away from Madrid to serve as decongestion zones.<sup>8</sup> Experience elsewhere shows that nearby overspill areas are much more successful than distant ones, and even then there are two major conditions for success: a very high level of shopping, social and cultural amenities and easy transport access to the metropolis by both public and private transport. Neither of these conditions holds in Spain. In practice, the 'poligonos' are little different from poles of development since their main scope for growth in the immediate future is by attracting population from the surrounding countryside rather than from the metropolis, and indeed the 'poligonos' have been established in areas suffering from out-migration.<sup>9</sup>

If the above arguments are correct, it follows that the level of inducements offered to decentralising firms is too low for the 'poligonos' to be successful. This is not a case for higher subsidies, but suggests that on cost-effectiveness grounds the scarce budgetary resources could be better employed elsewhere in the regional development programme. This does not necessarily mean an immediate abandonment of the 'poligonos de descongestion' policy, but rather that it should not be extended and that progress at the existing 'poligonos' should be continuously evaluated and reassessed with a view to

concentrating resources at the margin on those with the most potential.

### Conclusion

This paper has been highly critical of several aspects of Spanish regional policy: weak specification of policy objectives; the failure to relate policy decisions (e.g. the selection of poles) to goals; insufficient evaluation of instruments; inadequate co-ordination and co-operation among government departments concerned; the myopia of regional policymakers; the neglect of interdependence between regional industrial and physical planning. Nevertheless, to be offset against this, a regional policy has been in operation for only seven years, budgetary resources for regional development are scarce, regional planners in Spain are relatively inexperienced, they have to work in a very sensitive political environment in the central administration and in the regions, and the resources available for research and evaluation are limited. Moreover, there has been a marked convergence in regional *per capita* incomes in Spain in recent years though this owes more to the natural forces of inter-regional migration than to the effects of regional policy.

### REFERENCES

- BANCO DE BILBAO (1967). *Renta Nacional de España y su Distribución Provincial*.
- COMISARIA DEL PLAN. (1968). *Estructura y Servicios Urbanos*.
- (1968). *Incentives in Industrial Centers*.
- (1969). *Les Pôles de Développement*.
- (1969). *Análisis del Desarrollo Regional*.
- (1970). *III Plan de Desarrollo* (unpublished).
- COMMON MARKET BUSINESS REPORTS (1969). *Foreign Investments in Spain*. Geneva.
- INSTITUTO NACIONAL DE ESTADISTICA (1968). *Migration y Estructura Regional*.

<sup>8</sup> New 'poligonos' under consideration (Tarancon, Aranjuez and Talavera) are fortunately much closer to Madrid.

<sup>9</sup> Even so, growth of this kind is indirect decongestion. This is because the main alternative centre of attraction for migrants to the 'poligonos' is Madrid so that diverting them to the 'poligonos' helps to prevent a further build-up in the metropolis.

I.B.R.D. (1963.) *The Economic Development of Spain*.  
 LASUEN, J. R. (1962). Regional Income Inequalities and  
 the Problem of Growth in Spain. *Papers and Pro-  
 ceedings of the Regional Science Association*, Vol.  
 8.

— (1969). On Growth Poles. *Urban Studies*, Vol. 6,  
 February.  
 PRESIDENCIA DEL GOBIERNO (1969). *II Plan de  
 Desarrollo Economico y Social*.  
 — (1967). *Anexo al II Plan : Estudios Sectoriales*.

*Acknowledgement.* I am grateful to Dr Fernando Fernandez Rodriguez, Chief of the Regional  
 Studies Section, Comisaria del Plan, Madrid, and to Dr Friedrich Buttler of the Ibero-America  
 Institute at the University of Göttingen for discussions on Spanish regional policy, and to Sr Javier  
 Irastorza Revuelta, Secretary General of the Comisaria del Plan, for granting me working facilities.  
 I also wish to thank Mrs Eileen Mitchelhill of the Centre for Research in the Social Sciences, University  
 of Kent at Canterbury, for statistical assistance.

## 34 The Politics of French Regional Planning

George W. Ross and Stephen S. Cohen

Regional planning has become a necessity in most countries. But nobody seems to know quite what it is, and no nation seems to know how to do it. This paper analyzes France's twenty-five year experience of trying to fit a geographic dimension to its national planning and establish political and administrative structures appropriate to those regional objectives.

No nation in the Western world inherited such an overwhelming legacy of centralization as France. The triumph of the absolute monarchy in the sixteenth and seventeenth centuries over the feudal nobility concentrated French political, administrative, and cultural life at the court of Versailles and in Paris. The revolution and its consolidation under Napoleon gave France a new politics but reinforced the old centralization. The Jacobin state and the Napoleonic order were both Paris-centered to an extreme degree. Throughout the nineteenth and early twentieth centuries, the model did not change. Activity was concentrated in Paris—power, money, intellect, culture—while everything provincial was thought to be, and probably was, helplessly backward and largely irrelevant. Even the spread of capitalism, while leading to the development of certain provincial industrial centers, actually enhanced the predominance of Paris over the rest of the country. As modern communications networks were built, they all had Paris at their center. And, of course, economic centralization went along with, and in turn encouraged, centralization in the political, intellectual, and administrative spheres.

The monocentric society that developed in France was striking, compared with other industrial societies. In Germany and America political and economic decentralization went along roughly together. England was a different case, of course, but even there a large degree of local autonomy prevailed. London was not disproportionately overpowering (although it may have become more so since 1945), and the way in which industrialization developed led to powerful regional foci. In all of these countries polycentrism was a reality, and self-sufficient middle-sized cities emerged as the loci of a lively regional life and development even where huge urban centers existed. This was not true of France, where growth was marked by the absence of substantial middle-sized centers. In France, there was Paris and the provinces; everything began and ended in Paris.

In the immediate postwar years, a seminal book by Jean-François Gravier,

Research for this paper was supported by the Office of Economic Research, Economic Development Administration, U.S. Department of Commerce. © 1974 by George W. Ross and Stephen S. Cohen.

aply entitled *Paris and the French Desert*, brought the question of France's geographical imbalance before the reading public. The book was well-timed; current events made it necessary to take the question seriously. In the postwar period the French economy began rapidly to overcome its backwardness relative to other industrial societies, and the pattern of economic growth and social change that resulted greatly accentuated the country's territorial disequilibria. New sectors of economic life mushroomed in the Paris region while older sectors appeared to be dying in the provinces. Textiles and mining in the north and shipbuilding on the Atlantic coast, for example, were struck by what appeared to be an irreversible decadence. In short, the economic modernization of postwar France further upset the already precarious balance between Paris and the rest of the country.

Traditional hypercentralization plus the centralizing tendencies of postwar modernization were two factors leading to the recognition of the need for regional planning. The plight of France's agrarian sector was a third. Widespread small holdings, a peasant tradition and ideology, and agricultural protection had kept a large percentage of France's population on the land well into the twentieth century. Since small peasant plots had long since ceased to make much economic sense, the situation was inherently unstable. The predictable exodus off the land began in earnest after the war. While migrations from the countryside to the city helped the industrial labor market (although they contributed to urban difficulties), they were at the same time a sign of crises on the land. Modern agrarian technology meant a decline in rural agricultural employment. Where the problem was faced, serious rural depopulation threatened; where it was not faced, misery and decline followed. In many areas the young, ambitious, and qualified departed for the city—usually Paris—leaving the others behind. Those least prepared to change were thus left to confront a situation where change was obligatory. Non-agrarian jobs in rural areas, or even in urban areas close to stricken rural areas, would not exist in the absence of deliberate action. The general lack of industry in much of provincial France meant that if nothing was done to promote industrial decentralization and provincial economic development, whole regions might become wastelands while pressure on Paris could intensify to an intolerable point.<sup>1</sup>

Regional planning was also prompted by increasing evidence of a striking imbalance between west and east in the country. France west of a line drawn from Le Havre to Marseilles had much more than its share of both industrial and agricultural backwardness. Where industry existed in the west, it tended to be obsolete and dying (shipbuilding, fishing) with little moving in to replace it. Both agrarian and industrial sectors in the west were the victims of

<sup>1</sup>See Gordon Wright, *Rural Revolution in France* (Stanford, 1964); Serge Mallet, *Les paysans contre le passé* (Paris, 1964); J. R. Gravier, *Paris et le désert français* (Paris, 1947); Pierre Viot, *L'aménagement du territoire en France* (Cours, Institut d'études politiques, Paris, 1967), chap. II; R. C. Richard, *Regional Planning and Horticulture in France* (London, 1968).

poor communication with the rest of France—a problem which grew more and more critical with the opening up of new European markets to the east. Moreover, the west was both relatively underpopulated and highly agricultural. To prevent disaster something had to be done for the west beyond what the west seemed able to do for itself. This was also true of certain other areas of France where declining industry was dominant. In the north (the textiles areas of Lille-Roubaix-Tourcoing and the coal mining areas around those cities) and in Lorraine (iron mining and steelmaking, not in decline but in great transition) alternative employment for the large industrial labor force already present in these areas had to be created to prevent the development of huge regional slums. New industry had to be attracted to these areas, and retraining and readaptation programs had to be set up for the workers involved. Here major problems of industrial conversion had to be faced.

Finally, France's postwar experience with general economic planning itself tended to lead toward regional planning. The planners' experience revealed that maximal economic growth, the object of their endeavors, did not necessarily lead to a spatial economic optimum—a vague notion of regionalized equity. Indeed, it seemed that the accentuation of regional disparities which accompanied France's high growth rate would likely lead to bottlenecks and greater long-term problems. This knowledge prompted the planners to think in terms of planning in space as well as time, an innovation that would involve some entirely new and technically challenging priorities.

### Origins of Regional Planning

#### Early Planning

Promoting regional development had some precedents in French history. For example, in the interwar period, the aeronautical industry was moved, for strategic reasons, far away from the Germans and spread around Toulouse in the southwest. As is so often the case in economic history, military operations had unforeseen economic effects: Toulouse, now the center of the French aerospace industry, is one of the few dynamic cities in the southwest of France. The Toulouse experience, long forgotten by defense strategists, has recently been resurrected by regional planners in the form of growth pole models of regional economic development.<sup>2</sup> Another example of regional planning as a by-product can be found in the Vichy government's efforts to create, for its own very particular purposes, regional political and administrative structures—thereby discrediting a perfectly good idea for a considerable time. Such schemes, cleansed of their historical purposes and connotations, are also coming back to life.

But it was not until the post-World War II period that circumstances forced the authorities to face problems of regional development seriously. The administrative body that felt the most immediate effects of the nation's fast-growing geographic imbalance was the Ministry of Construction. Within

<sup>2</sup>See Bernard Pouget, *La délégation à l'aménagement du territoire et à l'action régionale* (Paris, 1968), part I.

the ministry, a *Département d'aménagement du territoire* (DAT—literally a department to “arrange” or “fix up” the territory) was formed under the guidance of Marius Claudius-Petit, a key figure in the early development of regional planning.<sup>3</sup> The point of departure of the DAT was “urbanism” (urban physical planning), the provision of housing and urban infrastructure.<sup>4</sup> At that time, following the neglect of the thirties and the destruction of the war, urban facilities were woefully inadequate, and the mass migration from farm to city, most often to Paris, threatened to lead to chaos. It was but a short step—at the level of economic theory—from accommodation to redirection. Pressures on Parisian urban facilities could be reduced if the flow of migrants could be channeled to provincial centers. And the key to location was jobs. Thus, for the Ministry of Construction, the creation of jobs in provincial centers, promoting industrial decentralization, became the aim of the rather crude and limited policy that emerged.

The Ministry of Construction attempted to shape the industrial geography of France by various “directing” techniques, most of which were negative. One important way to achieve industrial decentralization was to use state powers to block certain activities deemed undesirable from a regional planning point of view. The issuance of building permits was the key to this tactic. In the system of prior authorization that emerged from the ministry, plans for any new building of over 500 square meters floorspace, or for one designed to hold more than fifty employees, had to be submitted for approval. In principle, business that did not have to be in Paris would be denied construction permits. Rigorously applied, such a system could control the inflow of new plants to the Paris area. But how effectively this policy was applied, how much power the regional planners could actually exercise, was another question.

Negative control was not the only instrument of regional policy. Over these early years, administrators tailored a whole panoply of subsidies, loans, quick write-offs, and outright gifts to give themselves some positive powers of impulsion. But control of the most important of these positive aids remained securely in the hands of the Ministry of Finance, for which considerations of regional balance were just one priority among many. The Ministry of Construction had only limited resources, and the regional planners of the DAT had only limited influence within the Ministry of Construction. From the very beginning it was clear that successful implementation of any regional development strategy would depend upon a coordinated commitment of the principal governmental bodies, especially the most powerful, the Ministry of Finance. Major political forces had to be mobilized, administrative structures

<sup>3</sup>Ibid. See also Jean Lajugie, “Aménagement du territoire et développement économique régional en France (1945–65),” *Revue d'économie politique*, Jan. - Feb. 1964, part I; Marius Claudius-Petit, *Pour un plan national d'aménagement du territoire* (Ministère de la reconstruction, Paris, 1950).

<sup>4</sup>Pouget, p. 27.

formed, and a development strategy equal to the task and appropriate to the political context created.

This early form of "directive" regional planning had serious limitations, in addition to its obvious lack of clout. In a period when Jean Monnet and his group of national economic planners were slowly and painfully acquiring data, theory, and techniques for their task of "sectoral" planning, the regional planners, faced with a considerably more complex task, were almost totally without adequate information or theoretical perspectives. Moreover, Marius Claudius-Petit and his companions at the Ministry of Construction were essentially forced to respond to one rather urgent problem, that of urban equipment to accommodate immigrants, which limited their perspective considerably.

Perhaps the most prominent characteristic of these early efforts—aside from their weakness—was that they were basically top-down, bureaucratically devised and imposed from Paris, from the center. Even when the question was that of developing provincial France to balance Parisian centralization, it was Paris that dictated how this was to be done. This was regionalism from the center, with no decentralization of decision-making at all.<sup>5</sup>

After 1950 this centralized direction of decentralization ceased to have the field of regionalism to itself. Independent "committees for regional expansion" emerged locally (of which the Breton committee was perhaps the most vociferous and well organized).<sup>6</sup> No single factor explains why these committees came into existence when they did. In areas such as Brittany, ethnic bonds complemented economic grievances. In others, a growing awareness of existing economic backwardness and imbalance and of trends toward their worsening prompted the committees either to pressure Paris for more vigorous regional development or to attempt to promote development themselves. Most of the committees were in one way or another explicit protests against Paris. Weak regionalism from the center seemed to generate equally weak demands for grass-roots regionalism.

These early efforts at regional decentralization from the center on the one hand and the emergence of regional expansion committees on the other provided a foretaste of many of the basic problems in the later development of regional planning in France. From its beginnings regional planning was generally technocratic and centralized. The process was administrative, the immediate problems that the planners faced were either administrative (traditional bureaucratic habits and divisions of labor) or technical (lack of theory and data on regional development). Regional planning emerged within the special politics of economic administration, well outside the traditional

<sup>5</sup>Jean Chaffel, *Autocritique de la France* (Paris, 1965), p. 123.

<sup>6</sup>Michael Philliponeau, *La gauche et les régions*, chap. III; Charles Roig, "Les aspects socio-politiques de la planification régionale en France," *Economique* (Montreal), April-June 1967, section II.



political arena.<sup>7</sup> It involved no formal (and in its early stages little informal) consultation with political and social forces in the country at large, not to speak of participation of such forces in the decision-making processes. Little wonder, then, that the processes elicited regional responses such as the regional expansion committees. No doubt those responsible for regional planning had a tendency to see questions in technical, administrative terms. This, after all, is what technocrats do. But it must be pointed out that there was some wisdom, whether conscious or not, in such an approach. Those most active in the regional pressure groups, those who would most have desired to turn regional planning into a politicized process were traditional provincial *notables* of both the left and the right.<sup>8</sup> Their perspective was distributive, not developmental; they aimed at incorporating regional development into the traditional local political structures that they controlled, not at developing new ones. There seemed to be much to be gained in keeping such forces from having a significant voice in the planning process since they might have channeled it into the traditional scramble for handouts, with little regard for the objectives of the regional development policy and ignorant disdain for the more subtle matters of planned strategy, efficiency, and coherence. In any case, the outlines of many subsequent problems were set out. Was regional planning to be technocratic and centralized? How far could such technocratic centralized planning go? How could true regional participation in the planning process be devised without undermining the delicacy of the process itself? What form would such participation assume? A great deal of the recent history of regional planning revolves around such questions.

#### Development

With the 1954-55 Mendès-France and Edgar Faure governments, perspectives on regional planning began to change from urbanism and industrial decentralization to a much broader view of decentralized economic development. This change was accompanied by the beginning of a process of administrative reorganization to accommodate the new perspective and the parallel involvement of the Commissariat du plan in the regional planning picture alongside the Ministry of Construction, which had been the principal agent. Finally the period was characterized by a steady evolution of the techniques of implementing regional objectives available to the planners.

Experience up to 1954 had proven that the dimensions of regional imbalance were such that a haphazard and fragmented policy of industrial decentralization was insufficient to the task. It was discovered, for example, that reconverting decaying industries and areas (the textile area of the north, or the shipbuilding industry of Nantes and St. Nazaire) involved retraining and perhaps relocating the labor force, providing all sorts of strategic public

<sup>7</sup>See Stephen S. Cohen, "Planning as Politics," *American Economic Review*, May 1970, for a discussion of the political logic of apolitical planning, that is, planning within the special politics of economic administration.

<sup>8</sup>Roig, "Les aspects socio-politiques," section II.

equipment (such as schools, research establishments, and communications), amalgamating existing industries into efficient productive units, and so on. Likewise, in devising a strategy for the development of backward areas of the country, considerations of what kind of economic activity would be appropriate and most likely to lead to cumulative growth and what kind of support actions and facilities would be necessary to promote the success of such activities obviously went far beyond the sprinkling of branch plants that characterizes simple industrial decentralization. Regional planning began to evolve in the direction of a more comprehensive perspective, away from the piecemeal tactics of its immediate postwar origins.<sup>9</sup>

Ways had to be devised to assess the broader problems of regional imbalance and to administer the planning measures that were eventually proposed. For planning purposes France was divided up into twenty-one administrative regions. Several adjoining administrative departments were clustered into a region with a population of at least two million.<sup>10</sup> Any resurrection of ethnic or traditional regional entities was avoided. Administrative convenience was key. What these regions were to do was as yet unclear, except to serve as a transmission belt for problems and policies. Methods of solution in this period remained much the same as they had been under the Ministry of Construction: negative action to limit undesirable activities (by "prior authorization") and an assortment of positive encouragements to projects that seemed to merit promotion—subsidies, tax advantages, loans, and even buildings already constructed. Whole geographical areas (usually those where acute problems of reconverting dying industry existed) were designated "critical zones" to receive priority in the dispensation of encouragements.

During the same period the national Planning Commission began to move seriously into the area of regional planning; since the commission was not tied to the Ministry of Construction, the earliest promoter of regional development, two parallel and unconnected agencies were now in the field. The Second French National Plan had earlier expressed pious wishes about the desirability of regional development but had otherwise done little. Now the commission started in earnest to promote the preparation of "regional plans" for each of the new administrative regions. At both the national and the regional levels the primary aim of this process was pedagogical.<sup>11</sup> Regional administrative authorities and regional economic and social forces (specifically the autonomously developed committees for regional expansion) were to be led by participation in the planning process to "think regionally" (since the new regions were still paper categories with which no one had any

<sup>9</sup>See Lajugie, "Aménagement du territoire," part II, and Viot, chap. II.

<sup>10</sup>See decrees of 30 June 1955, 28 October 1956, and 2 June 1960.

<sup>11</sup>Pouget, chap. I, section II. See also Bernard Pouget and Patrice de Monbrison-Fouchère, "La régionalisation dans le IV<sup>e</sup> plan: l'expérience des tranches opératoires," in *Cahiers de la fondation nationale des sciences politiques*, no. 135, *Administration traditionnelle et planification régionale*, pp. 153–158.

practical familiarity) and to reflect on their region's needs and role in the national economy. At the other end of the process, the Parisian planners themselves hoped to learn a great deal more than anyone yet knew about the economic and social reality of the "provinces," in order to develop a coherent approach to regional planning.

The regional plans thus produced (the first was for Brittany, in 1956; others were slowly produced in the years that followed, but only thirteen of twenty-one regions had been done by 1964) were part training exercise and part information gathering. Generally they contained detailed inventories of the economic and social "givens" of the regions plus lists of economic objectives and possible projects for regional development. The perspectives of the plans were long-term and were not synchronized with the five-year national plans. Moreover, the projects thought desirable were just listed, not ranked or phased. It was obvious that they were essentially inoperative as plans. They were not meant to function as guides to action.<sup>12</sup> The Ministry of Construction continued to follow its own criteria in carrying on whatever practical regional development activities existed.

Producing plans for each region might have been a significant step toward enlisting local participation in the planning operation and thus decentralizing to some extent the administration of regional development. Indeed, claims were made that this was a central aim of the project. Whatever the intent, little or nothing was accomplished in this direction. Even in administrative terms the operation was centralized: the regional plans were largely the product of Parisian civil servants.<sup>13</sup> Politically, local participation of non-governmental groups was not extensive. There were several reasons for these facts. First, the task of projecting regional development posed very great technical problems for the central administration that local groups could not help. Moreover, it was quickly discovered that local groups lacked the technical competence to keep up with the administrators, despite the problems the administrators themselves had.<sup>14</sup> Finally, the "region" existed only on paper, and local groups had great difficulty in adjusting to it. They focused instead on departments or on more traditional and more functional regional identification.

Thus, as French regional planning expanded, the perspective of "reform from above and from the center" returned. Once again it appeared that regional development was to be bureaucratically and centrally decreed. It was not surprising, then, that the resentment of local groups and old regional elites grew, especially as regional development from the center was doing very little to redress the ever worsening regional imbalances.<sup>15</sup>

<sup>12</sup>Jean Milhau, "Le financement de l'expansion régionale," *Revue d'économie politique*, Jan.-Feb. 1964, p. 115.

<sup>13</sup>Pouget and Monbrison-Fouchère, pp. 156-167; Chaffel, p. 123; J. de Lanversin, *L'aménagement du territoire* (Paris, 1965), pp. 14, 33.

<sup>14</sup>See Bruno Vennin, "Aménagement du territoire et initiative démocratique," *Esprit*, Feb. 1967, p. 348.

<sup>15</sup>See Georges Lavau, "Aspects politiques du problème des économies régionales," *Formation*, Sept.-Oct. 1964.

By the late 1950s it was possible to assess the results of the program of industrial decentralization and regionalism promoted essentially by the Ministry of Construction. Undoubtedly some success at industrial decentralization had been achieved; industrial activities that would have gone to Paris had been sprinkled here and there. But to cite Jean Lajugie, "the experience of ten years ought to make apparent the need of global action. The initiatives taken revealed themselves too often to be dispersed, incomplete; the responsible organs too often lacked means of action commensurate with their good will . . . it is too evident that political will was lacking during this period, likewise lacking was the budgetary continuity indispensable to the execution of the chosen solutions. . . ." <sup>16</sup> Nearly all observers agreed that much of the industry that had been dissuaded from locating in the immediate Paris region had then proceeded to locate around the capital. <sup>17</sup> Furthermore, this decentralization was primarily confined to production and assembly plants whose administrative, clerical, and research facilities remained in Paris. <sup>18</sup> Such a pattern did little to alter the general economic imbalance of the nation; it merely helped to spread out Paris. Despite the program for industrial decentralization, the Paris region was growing by leaps and bounds. It had become a true urban-suburban mess with accelerating problems of housing, transportation, services, and so forth. Moreover, it was fast becoming the office and brain for all of France.

Inherent defects in the whole industrial decentralization approach to regional planning had become clear. The blow-by-blow approach led to a generally incoherent sprinkling (the French commentators called it *souppoudrage*, "powdering") of new industry here and there under anything but the best circumstances. (Lajugie speaks of the location of firms "encouraged without sufficient study of possible markets" whose "fragility" constituted a "permanent peril" to localities "fortunate" enough to get them.) <sup>19</sup> Little in this approach guaranteed a cumulative process of economic development. An electronics assembly plant transferred to the Loire Valley to take advantage of cheap peasant labor might be an example of industrial decentralization and might provide jobs for a certain number of women, but it would not initiate a process of cumulative economic change in the area. This was especially true if the firm kept all of its administrative and research facilities in Paris, as most did. Likewise, the relocation of relatively marginal firms in new areas where they were given a new lease on life by civic support, state subsidies, and a supply of docile, cheap, and mostly female labor held very little promise of leading to a process of cumulative development. <sup>20</sup> Finally, even when

<sup>16</sup>Lajugie, "Aménagement du territoire," p. 289.

<sup>17</sup>Viot, chap. I; Lajugie, "Aménagement du territoire," p. 304; Lanversin, p. 363; R. Minguet, *Paris et l'hexagone française* (1962), p. 25.

<sup>18</sup>Serge Mallet, *La nouvelle classe ouvrière* (Paris, 1963), has an interesting monographic discussion of the "decentralization" of the then French Compagnie des machines bull. See his chapter "Une usine deconcentrée, La compagnie des machines bull," pp. 77-141. See also Lajugie, "Aménagement du territoire," pp. 302-304, 321-325.

<sup>19</sup>Lajugie, "Aménagement du territoire," p. 385.

<sup>20</sup>With reference to local political economies, see Charles Roig, "L'administration traditionnelle devant les changements sociaux," *Administration traditionnelle et planification ré-*

industrial decentralization on a broader scale was successful there still seemed to be no guarantee that economic and social development of a magnitude sufficient to begin redressing regional imbalance would follow. It was clear, then, that merely sprinkling branch plants and marginal businesses over the map was not the answer, even if it were done in a systematic way—which it had not been.

On a purely administrative level, the emergence of both the Ministry of Construction and the Commissariat du plan as regional planning agencies was obviously a problem. The Ministry of Construction held many levers of power but was interested primarily in urban problems and in limited "regional action," while the Planning Commission was developing a much broader conception of development but had less influence over the actual mechanisms for implementing objectives and no experience with the actual practice of regional development. At the same time, however, the Ministry of Construction had demonstrated important weaknesses in its handling of the question.<sup>21</sup> For example, it had been relatively unable to harmonize the approaches of the various national ministries around the objectives of regional action, something that had to be done if any kind of coherent planning was to be undertaken. The DAT tended to plan "projects"; it never planned for a planning process—an institutionalized political force capable of affecting the basic economic and political forces that determine location. Moreover, the resistance of local levels of the administration to regional objectives remained strong.

The duality of planning agencies and of views on planning made little sense. Several things indicated that the Planning Commission was the more appropriate agency. Its conception of the problem was broad and comprehensive, while the construction ministry's narrower approach had been shown insufficient by experience. Most important, the Planning Commission had shown in its experience with general economic planning an ability to create a harmonization of objectives and practices among traditional ministries, which was the key to the regional planning task. Indeed, the Planning Commission had been held up as a model agency for getting around the bottlenecks created by traditional administrative demarcations and habits. Finally, the commission had also shown a particular inventiveness in devising functional new institutions to achieve new tasks.<sup>22</sup>

#### Progress during the Fifth Republic

When General de Gaulle came to power in 1958, regional plans were being produced, although very slowly and more for educative than for immediately practical purposes.<sup>23</sup> They served essentially to get central administrators to

*gionale*, p. 65, and Mallet, *Les paysans*.

<sup>21</sup>Roig, "L'administration traditionnelle," p. 36; Jean-Louis Quermonne, "Planification régionale et réforme administrative," in *Administration traditionnelle et planification régionale*.

<sup>22</sup>See Stephen S. Cohen, *Modern Capitalist Planning* (Cambridge, Mass., 1970), for an analysis of French national planning.

<sup>23</sup>Quermonne, "Planification régionale," p. 92.

think regionally and to begin gathering the most basic kinds of information about the "regions." Little was known about regional economic structures and even less about such things as interregional economic relations and complementarities. Better data were becoming available about demographic and manpower trends, but even these were not yet in a form useful for regional planning. Without such knowledge, of course, little could be accomplished, as ten years of attempts to work without it demonstrated. One key problem of regional planning was intellectual: with the best will in the world relatively uninformed programs of industrial decentralization had left much to be desired and had often led nowhere. On the other hand no workable model for operational regional planning existed.<sup>24</sup>

By the early 1960s, in the administrative realm, with DAT elaborating its own *plan vert* and the Planning Commission working on the Fourth Plan, the situation appeared absurd. To quote Bernard Pouget, "the simultaneous elaboration of these two plans without harmonization in their lengths, their geographical framework or in their major propositions, had at least the merit of drawing the government's attention to the necessity of establishing a coordination between the DAT and the Planning Commission."<sup>25</sup>

The question of centralized versus locally based regional planning remained open. Ten years of tentative regional planning had been highly centralized and administrative. It appeared that regional development meant more centralized bureaucratic power. The reasons for this were obvious. The "region" was as yet a paper entity politically and administratively. The foci of French politics were either local or national. Local politics was largely the affair of the *notable*—notorious for his narrowness and conservatism (right or left). Any politicization of regional planning that created a large role for unreconstructed *notables* would risk turning the planning process into a forum for the pettiest local interests at the expense of rationality and coherence. The administrators were interested in regionalizing national development, not in regional problems per se, and they felt this national and developmental perspective, which they considered primary, might be risked by premature politicization.

One other issue was critical. The administration itself focused only nationally and locally; no established regional level of administration as yet existed. Only the national administration thought regionally. The focus of local administrators was departmental. One simply could not trust administrators with a purely local focus, caught up in traditional local systems of power and influence, to carry through regional planning, for many of the same reasons that one could not trust regional planning to local politicians. The construction of a regional administrative echelon was clearly in order.

Thus the agenda for French regional planning was clear. Information had to be gathered, new approaches and techniques developed. The strategic

<sup>24</sup>See Viot.

<sup>25</sup>Pouget, p. 32.

focus of planning had to be clarified, and a method of carrying out regional planning priorities had to be established. For all this, a regional administrative structure was necessary. It seemed certain, then, that major administrative reforms would have to be the next step. Even the concrete results of regional planning would for a time be secondary to this primary task of building a satisfactory administrative context.

#### National Administrative Change

The division of regional planning activity between the Ministry of Construction and the Planning Commission was gradually resolved in favor of the "plan," the turning point being the early 1960s. This resolution was of immense importance in terms of the perspective within which regional planning was to be carried on. The Plan was a national institution, a complex of activities to promote long-term national expansion and development. This was a point of view that could lend itself to regional development, but in a specific way. As the planners moved into regional planning it became apparent that they were not primarily interested in stimulating the emergence of new territorial entities, nor in planning regional economies as such. Instead, what they wanted was to give regional coloring to their ongoing national planning programs; the whole process would involve not greater regional autonomy but greater regional integration into the national decision-making operation.<sup>26</sup>

The job of regionalizing the Plan fell to the Planning Commission. Nine sectoral specialists in regional problems were added to the lightweight staff of the Plan, and its sectoral commissions were prodded to consider the regional aspects of each sector's development.<sup>27</sup>

Implementation, primarily on the level of national government but also on the regional level, was equally rudimentary. Here the real question was how to get several tradition-bound, autonomous, and functionally specific administrative bodies, terribly jealous of their prerogatives, to think and act regionally in terms of the planners' aims. Coordinating the activities of such staid bodies was difficult enough in itself, but coordinating them to do things in a very new way according to priorities that they themselves did not play a primary role in setting was a huge task. Fortunately, precedents and something of a model for dealing with such problems existed. The National Plan had grown up in similar circumstances and had worked out passably well. Thus it was no surprise that the new body to regionalize the priorities of central administrations by getting around the classical administrative structures was modeled on the Commissariat général du plan.<sup>28</sup> What was needed

<sup>26</sup>In general, see Jean Lajugie, "La politique française de développement économique régionale de 1958 à 1963," *Revue juridique et économique du sud-ouest*, no. 2 (1963); idem, "La planification régionale," *Revue juridique et économique du sud-ouest*, no. 4 (1966); Jean Milhau, "La politique de l'expansion régionale en France," *Revue des sciences économiques* (Liege), Dec. 1963; Lavau.

<sup>27</sup>See Commissariat général du plan, *Commission générale de l'aménagement du territoire, premier rapport* (Paris, 1964), and *Premier rapport de la CNAT* (Paris, 1964); Viot, part II.

<sup>28</sup>See Cohen, *Modern Capitalist Planning*.

at the top to work administrative change was a "light" organization (flexible and small)—not large enough to do anything on its own and therefore forced to work through other bodies to achieve its objectives—staffed with high-powered polyvalent administrative types, to serve as an honest broker in animating various ministries to regionalize according to the aims of the Plan. Such a light organization could move around the administrative structure to organize influence and intervene in the preparations for administrative decision-making to promote regional actions.

The Délégation à l'aménagement du territoire et à l'action régionale was the body finally set up in 1963 to do the job. At the head of DATAR was a minister, a sign of the importance assigned to the body. The man holding the post through most of DATAR's existence was Olivier Guichard, a very important Gaullist political strategist. Under Guichard a small, high-powered staff with a strong interdisciplinary mix was gathered. DATAR had ambitious objectives. It was to coordinate regional development activities on a national level, on a local level, and between levels. It was to be the general synthesizing organization for regional planning, making sure that things got put together, that those who ought to know things did know them, and that all concerned were well disposed toward the planners' regional goals. It was charged with establishing a workable process of arbitration between ministries on regional issues. In this respect it served as the secretariat of an interministerial Committee on Regional Planning, a position that gave it some leverage. Finally, it was to serve as a general center of animation for regional planning, promoting activity to keep the planning and execution program thriving wherever such activity was needed. More specifically, DATAR was to play a key role in one of the central processes of regional planning, the regionalization of the budget for public investment, which involved activity on a national level and on a regional level and between them. DATAR also played an important role in water resources policy, industrial decentralization, urban planning (with the great exception of Paris), developing national parks, and so on. In all of these areas DATAR had specific powers of its own, plus influence over more general systems of incentives to see that regional planning was carried on.<sup>29</sup>

It also had a relatively small allotment of money under its own control, Fonds interministeriel pour l'aménagement du territoire en France (FIAT), about \$50 million per year.<sup>30</sup> FIAT resembled DATAR—by design. It was too small and too weak to do anything on its own. It would have to operate

<sup>29</sup>See Pouget for more complete description; also see Jean-Louis Quermonne, "La régionalisation et l'aménagement du territoire en France," in Institut d'études politiques, Grenoble ed., *L'aménagement du territoire et le développement régional: 1965-1966* (Paris, 1968). This is intended to become an annual manual on such questions with several articles of great use, plus charts, an exhaustive bibliography, and reproductions of important legislative texts relating to regional planning.

<sup>30</sup>See Organization for European Cooperation and Development (OECD), *The Regional Factor in Economic Development* (Paris, 1970), p. 66, for FIAT budget.



in close conjunction with established sources, supplementing, stimulating, demonstrating.

DATAR, modeled after the Planning Commission, was deliberately designed not to pose a direct threat to existing ministries. The fact that it was too small in staff and resources and too weak in legal powers to act on its own constituted a fundamental guarantee to the ministries: like the Plan, it cannot replace them; it cannot command them. DATAR cannot become a super-ministry. It must work within the existing structures of bureaucratic competence and power, trying to initiate and coordinate action by other ministries. But unlike the Plan, which developed during a period of weak, unstable governments and strong, independent bureaucracies, and consequently stressed political noncommitment and independence, DATAR was created in a period of strong, Gaullist governments. It has been much closer to purely political undertakings than the Plan was in its early days. Though headed by a "minister," DATAR is not an independent ministry. It is attached directly to the Prime Minister's office. The Plan has also been attached to the Prime Minister's office at various times in its life. But DATAR's attachment has been more intimate in political terms. Among its other consequences, this close political attachment has been an important source of its influence in dealing with other administrations.<sup>31</sup>

#### **Regional-Level Administrative Structures**

Prior to the reforms of the Fifth Republic, there existed no regional administrative structures appropriate to the task of regional development. Thus, from a purely mechanical point of view, to get the job done it was necessary to work administrative change. It is well to keep in mind, however, the nature of the "regional planning" we are dealing with. What was desired was a regionalization of French national planning—regional planning from the center, not regional planning from the region, and certainly not planning from any regional grass roots. The creation of a regional echelon of administrative authority for the purposes of "regional planning" meant an attempt to improve horizontal coordination on the regional level of branches of the national administration. This was not to be seen as any decentralization of power from center to periphery. It was not a transfer of power from the administration to other groups, but rather a movement within the administration, a "deconcentration" (as French commentators call it) of central administrative power, which was hoped would make it more sensitive to issues raised at the periphery and thereby more able to make geographically sensible national plans.<sup>32</sup>

<sup>31</sup>We are indebted to Charles Goldfinger's analysis of this and many other points in this paper. His forthcoming Ph.D. dissertation on French regional planning will treat the subject in the depth it deserves. See also Pouget, p. 37 ff., for a more formal description of the place of DATAR in the French administration.

<sup>32</sup>Pierre Grémion, "Résistance au changement de l'administration territoriale: le cas des institutions régionales," *Sociologie du travail*, July-Sept. 1966, p. 280; François de Baecque, "Pour une politique cohérente de déconcentration," *Revue française de science politique*, Feb. 1967; J. F. Gravier, *La question régionale* (Paris, 1970), p. 68.

Perhaps the most important change at the regional level was the creation in 1964 of the regional prefect.<sup>33</sup> For the purposes of regional planning the age-old practice of equality among prefects was modified. The regional prefect was charged with specific regional planning duties and given powers to carry them out. He was given certain rights of "consultation" and "information" on broader questions of the National Plan as they related to his region but has, in principle, real power only over the planning and execution of certain categories of public investment are specifically regional in nature. His main task is to elaborate and report on the execution of what is called the *tranche régionale* (literally, "regional slice") of the plan for his region. In order to do this he has to gather information on regional needs and trends from available sources, then coordinate the consultation and action of departmental prefects and administrative departments. His main foci for the planning task are regional manpower (one aim of this level of regional planning being the control of interregional migration) and regional public investments. Deliberately excluded from his regional planning purview, among other things, are taxation, higher education, enforcement of labor law, and the gathering of statistics. He is assisted in his task by a staff of ten to fifteen specially detailed civil servants, called the regional mission. He works in conjunction with a regional administrative conference composed of important civil servants, other prefects, and regional functionaries of various national ministries.<sup>34</sup>

The closest the regional planning apparatus officially gets to contact with the people and/or interests of the region concerned comes when the regional prefect fulfills his legal duty to "consult" with another new regional institution, the Commission on Regional Economic Development (CODER) on regional aspects of the National Plan and on the regional slice for which the prefect is personally responsible. Prior to 1964, the regional prefect consulted with the committees for regional expansion, which it may be recalled were bodies that emerged autonomously in certain regions. It may well be that these earlier bodies, because of their regional and/or local perspectives, threatened to interfere with the more national aims of the planners. In any case, the CODERs are bodies of twenty to fifty members, one-quarter of whom are local politicians (mayors, members of general and municipal councils, and deputies if they hold such posts in addition to their national mandates), one-half of whom are designated by "corporate interests" (agriculture, business, labor, and so on), and one-quarter of whom are people appointed, for whatever reasons, by the Prime Minister.<sup>35</sup>

What actually happens at the "regional" levels of French regional plan-

<sup>33</sup>Decrees 64-250-252, 14 March 1964.

<sup>34</sup>See L. T. Sweetman, "Prefects and Planning," *Public Administration*, no. 43 (1965); Quermonne, "La régionalisation"; decree 64-251, 14 March 1964; Gravier, *La question régionale*, pp. 69-72.

<sup>35</sup>"L'aménagement du territoire en France," *Notes et études documentaires*, no. 3461 (Feb. 1968), pp. 50-51; decree 64-252.

ning? What do the institutions described above actually do? The Planning Commission sets broad regional development objectives as part of the National Economic Plan. That the basic choices of regional objectives are made on the national level is, of course, quite consistent with the true nature of French regional planning, which, as has been pointed out, is not the planning of regional development in the usual sense at all, but the "regionalization" of the National Economic Plan. Once these basic choices are outlined, documentation and information bearing on the choices to be made at the regional level are sent to the regional prefects.

It is important to be clear at this point just what is involved in the preparation of the *tranche régionale*, just what choices are actually made at regional levels. The *tranche régionale* is essentially a five-year program of public investments within the region. Major national programs of public investments are outlined by the planners on a national level—things like major communications and transportation systems, university expansion, and energy policies. Thus the critical choices for regional planning are decided centrally; they do not figure as parts of the *tranches régionales* and are not within the purview of the regional prefect.

A second category of public expenditure—projects that are purely departmental in nature—also falls outside the purview of the regional-level planning team. (The map of France is divided into ninety departments, which constitute the basic administrative units.)

A third category of investments of a specifically regional nature, those not "individualized" at the national level, are the specific matter upon which the regional planners work. The totals of these investments are set and apportioned among the regions by the central body but, unlike the first category of investments, these have not been allocated to specific projects. Since they are essentially intraregional public investments rather than projects whose importance transcends regional boundaries, they are given to the regional planners to allocate. Included in this category might be housing projects, secondary educational establishments, and agricultural schools.<sup>36</sup> It is, then, the task of the regional planners to organize the use of these specifically regional funds for their region, and this task is the essential one in the preparation of the *tranche régionale*.

Generally, the amounts of public investment to be dealt with by a region in the preparation of its *tranche régionale* are quite narrowly specified in total by the various national ministries as a result of prior activity on the national level. This means that the regional prefect and his team receive from the top a series of departmental "envelopes" indicating how much public expenditure for activities falling under each ministry has been projected for their region, that is, how much for housing, how much for schools, how much for roads, and so forth. It is important to understand that the prefect does not receive a block grant. He cannot spend the road budget on universities, nor the

<sup>36</sup>Pouget and Monbrison-Fouchère, p. 164; decrees 64-250-252.

university budget on telephones. Nor does he ultimately carry out the approved projects. All activities are built and run by the appropriate vertical ministries. Preparing the *tranche régionale* involves trying to foresee how these allotted sums (which may or may not be rigidly limited) ought to be spent within the region over five years to best promote regional development. To do this it is necessary first to gather information, some of which may come from national services (the Plan, the national economic statistics gathering office, the ministries), from local sources (local manpower statistics, data on public spending), or from special studies commissioned by or carried out by the regional planning team's members (the regional prefect and his mission, the administrative conferences). Then the planners consult with various administrative departments and other competent groups within the region on what might be done and how it might be coordinated; much of this is done by the regional administrative conferences.

This final synthesis of information and priorities is the *tranche régionale*, a capital budget for intraregional investment over five years. The TR is subsequently presented by the regional prefect to the CODER for its "consultation." After this the TR is sent back up to the national planners, who adjust the totality of regional plans. Then the TR, equilibrated with nineteen others by the national planners, is submitted to DATAR, which proceeds to a regionalization of the budget, which is in turn submitted for discussion to Parliament.<sup>37</sup>

#### Assessment

Quantitative assessment of results is absolutely impossible. The primary aim of the entire series of institutional innovations on the regional level has been pedagogical and psychological rather than operational. What is desired is to promote "regional" thinking and the emergence of a "regional" focus for at least some form of decision-making, however dependent such a form may be on the national level. The strategy for achieving this is to create first of all a viable regional-level administration. Presumably if this were successfully accomplished, nonadministrative (political and social) forces would then turn toward the new level of decision-making and deliberation to wield their influence. The purpose, then, of the new regional structures is not to impose any new concrete model for regional development but to reorganize processes of influence.

Several problems have already arisen in the modest evolution of France's regional structures. First of all, and perhaps most important, serious questions have been raised as to whether the twenty-one regions created in the 1950s and still used make much sense as economic and geographical units. It has been clear for some time that a smaller number of much larger regional entities would provide a better approximation of the country's functional

<sup>37</sup>See Viot, chap. VI; J. R. Boudeville, *Problems of Regional Economic Planning* (Edinburgh, 1966); J. Marczewski, "Planification et aménagement du territoire (Cours de doctorat, Faculté de Droit, Paris, 1966); Monbrison-Fouchère, p. 189; Y. Prats, "La régionalisation du budget," *Revue de science financière*, Oct.-Dec. 1967.

economic regions. But the twenty-one regions were not designated according to simple developmental criteria. Administrative convenience figured prominently in their original delineation, and perhaps most important, a small number of large regions was anathema to true Gaullists, who felt (especially in the context of the Common Market) that they smelled of federalism, regional independence, and supranationalism.<sup>38</sup>

Whatever is decided about the rationality of the regions in an economic context, they have proved to be somewhat uncertain administrative organs, even in the terms that are set for them. The regional prefect, for example, the key to the new structures, has met considerable resistance in his new planning functions from other prefects in his region. As noted above, the creation of a superior rank of prefects for regional planning purposes ran against the French tradition of equality among prefects. In certain instances of conflict, departmental prefects seem to have sought redress from the regional prefect's power by going around him to Paris ministries to settle things. In other instances, departmental prefects have sought to push prefectorial unanimity on regional questions as a way of undermining the regional prefect's power.<sup>39</sup> The regional planning team headed by the regional prefect had, in certain circumstances, a tendency to function by bargaining with local elites, thus subverting their own supposedly directional role in the planning process. As a result, a coherent view of regional priorities became difficult to obtain. The regional prefect's administrative right arm, the regional mission, also had trouble in establishing an operational model for cooperative work among its members and tended to be isolated within the regional administration. Furthermore, the regional services of national ministries seemed to have been "regionalized" only slightly; instead they seem to have protected all of their prerogatives and autonomy against the regional planners.<sup>40</sup> In short, if the new regional institutions were supposed to have created a new regional administrative force, they have not yet clearly succeeded in so doing. The whole local administrative apparatus seems instead to be balancing between the traditional departmental and the new regional focus, not unnaturally hedging its bets on the success of "regionalization."

One thing that could definitely be asserted about the new structures was that they had had very little real contact with the "people" of the regions, however the "people" was defined.<sup>41</sup> The change from the participation of the committees for regional expansion (relatively autonomous homegrown bodies) in the preparation of the Fourth Plan to the "consultation" of the new CODERs for the Fifth Plan was seen in some sectors as a direct attack on

<sup>38</sup>See Michel Debré, *Au service de la nation* (Paris, 1963), p. 225. (Charles Goldfinger led us to this reference.)

<sup>39</sup>Grémion, p. 284; Roig, "Administration traditionnelle"; Jean Claude Groshens, "L'insertion de la région dans l'organisation politique et administrative," *Cahiers de l'institut français des sciences administratives*, no. 1, (1967).

<sup>40</sup>Grémion, pp. 288-291.

<sup>41</sup>See Serge Antoine, "L'aménagement du territoire," *Sondages*, nos. 3-4 (1966).

the burgeoning grass-roots regional movement in France. According to Michel Philliponeau, in Brittany, an area where popular regionalism was quite well developed, this change worked very effectively to set back the Breton regional organization.<sup>42</sup> The creation of the CODER was thus seen by at least some local and regional elites as a new device for the centralization of power rather than as anything regional.<sup>43</sup> The new institutions were designed to drain power away from local centers toward the center, not to transfer power from a local to a regional level.<sup>44</sup>

The practical operations of the CODERs and the earlier regional committees for expansion created further problems. The composition of the CODERs was entirely predictable, as were reactions to it. Examining the Loire CODER, Jean-Luc Bodiguel found a heavy overrepresentation of the "old" Right and Center—the hard core of old *notables*—with little left-wing or Gaullist representation, the whole reproducing the composition of the local *conseils généraux*.<sup>45</sup> In general, trade unions and labor were highly underrepresented or not represented at all. One reason for this underrepresentation, but only one among many, was that the labor movement, as in the case of the Loire, had decided to boycott the CODER in protest against what they saw as bureaucratic centralization undermining local democracy which might lead to some form of corporatism.<sup>46</sup> Composition was only one problem, however. It was felt by many that the CODER was a futile enterprise in itself, something of a sham to rubber-stamp administrative decision-making. Even where the CODERs were "consulted" in a serious way, they often found themselves confronted with elaborate syntheses already worked out by the staff of the regional prefects and permitted too little time even to understand the questions involved and too little margin for decision even if time had been available. In some cases CODERs (and committees for regional expansion) were virtually ignored.<sup>47</sup>

Thus the response of some local interests—demands for the politicization of the entire process—was understandable. In design and in practice, the forms of regional planning evolved in France did little to create a sense of participation on the part of anyone but the administrators. Whatever else was accomplished, this remained a serious problem.

### Objectives and Means: An Overview

Administrative change at national and regional levels has not been the only development in French regional planning in recent years. Significant new

<sup>42</sup>Philliponeau.

<sup>43</sup>Raoul Browne, "Du neuf en aménagement du territoire," *Perspectives*, 16 Dec. 1967, p. 13.

<sup>44</sup>Chaffel, p. 125.

<sup>45</sup>Jean-Luc Bodiguel, "Les commissions de développement économique régional," *Revue française de science politique*, June 1966, p. 484.

<sup>46</sup>*Ibid.*, p. 485. See also Pierre Lebrun, *Questions actuelles du syndicalisme* (Paris, 1964), Centre d'études socialistes, *La participation des syndicats dans la planification* (Paris, 1965).

<sup>47</sup>Pouget and Monbrison-Fouchère, part III; Chaffel; Grémion.

directions in strategy for regional development have emerged as well.

Attention shifted to the context of investment rather than the bare fact of investment itself. The proper context would attract investment that would be cumulative and region-building. Without this context investment would be wasted; it would fail to generate development. In other words, the planners realized that few private investors would forego the many advantages, economic, social, and cultural, of Paris; these were important even to those who were forced by industrial decentralization policies to build on a ring a hundred miles around Paris unless similar or compensatory advantages seemed available elsewhere.

Thus, despite the absence of a scientific theory of regional development, the planners began to work out a new strategy. The key to balanced growth lay in creating urban centers that would exercise a serious counterattraction to Paris. The core element of this strategy was to develop existing regional centers into "countervailing metropolises." Eight were projected. These countervailing metropolises were to be developed to provide prospective private investors with sufficient real attraction to take the heat off Paris.<sup>48</sup> How to do this was the question, and the answers varied with the city. Each metropole had to become a center for real decision-making instead of an industrial tributary of Paris. Tertiary activities which were incredibly centralized in Paris had to be decentralized. Communications and transport had to be adequate to a real city. Substantial regional elites had to be developed. Cultural and intellectual life had to be promoted. While it was obviously not possible to turn each metropole into a mini-Paris, it would be possible to promote an urban division of labor so that each would be able to rival Paris in some way, becoming a magnet for activity in some critical area.

The national planners thus sought to begin regionalizing an infrastructure of public investments so as to promote the emergence of these metropolises as serious alternatives to Paris for private investment. Communications among Lyon, St. Etienne, and Grenoble were improved. A new airport, to connect the complex to the outside world, was planned. The port of Marseilles-Aix was expanded and modernized, notably by DATAR's pet project at the Gulf of Fos along a "Rotterdam of the Mediterranean" theme. Nancy/Metz/Thionville were tied together by new roads, and so on. Attempts were begun to push avant-garde economic activities over which the state had some control in the direction of cities that had already shown sure aptitude to welcome them. Lyon-St. Etienne-Grenoble was promoted as an electronics center; Toulouse as the center of France's growing aerospace complex. Efforts were begun to expand or create research and development facilities, often as parts of university decentralization, to complement such activities. Indeed, one of the general goals of the new strategy was, to cite Pierre Viot, "to make the

<sup>48</sup>"L'aménagement du territoire en France," p. 30. See also Pierre Georges, "Les villes métropoles," in *L'aménagement du territoire et le développement régionale: 1965-1966* (Paris, 1968).

University one of the bases of the development of the metropole by the multiplication of contacts between University and industry, by the development of university research."<sup>49</sup> Although this goal was often partially blocked by the reluctance of certain intellectual establishments to leave Paris (notably the *grandes écoles*) some success was achieved, primarily through the creation and growth of new regional facilities but also through some real geographic redistribution as well.

Governmental agencies themselves tried some decentralization: Electricité de France transferred its nuclear research facilities to Marseilles, and non-directive operations of the Ministry of Education and the Office of Tax Collection were "regionalized." Efforts to promote the decentralization of certain tertiary activities over which the state had some control—banks, insurance, and other financial activities—were seriously begun. In brief, the planners wanted the metropolises to have economic infrastructures that would give them a fighting chance to attract private capital, to be endowed with intellectual and cultural facilities of their own, and to have tertiary activities pushed in their direction. All proceeded as if the planners really meant business, the business being to furnish France with the vital regional centers whose absence had long distinguished France's development from that of other industrial societies.

Efforts were made to move beyond hit-and-miss use of state aids and subsidies to investment in the direction of a more regionally rational distribution. To do this the country was divided into five zones, each ranked on an order of priorities for state aid to private investment. In the cities of the underdeveloped west, maximum subsidies for new investment and expansion are almost automatic—and significant since they can amount to 20 percent of total investment. The second zone is the "conversion" zone, areas in which declining industry unreplaced by new activity has created problems (the textile and mining areas of the north, for example). Here large grants for industrial adaptation are available. The third zone includes slow-growing towns in the more affluent areas of France that are eligible only for limited fiscal aids. The fourth zone includes all of the wealthier part of France except the Paris region and in this zone nothing can be assumed about state aid, individual evaluation being made of specific projects to determine eligibility for subsidies. The fifth zone, the Paris region, gets no special advantages.<sup>50</sup> It is interesting to note that this division of the territory for purposes of operational economic policy seems to ignore the twenty-one planning regions.

These newly evolved general priorities essentially reflect two different vague "models": one, the urban backbone model of economic polarization around new urban foci; the other, a system of more homogeneous regions: the underdeveloped, slow-growing west to be stimulated, the relatively prosperous east to be supervised, and Paris to be kept under control.

<sup>49</sup>Viot, p. 153.

<sup>50</sup>Boudeville, pp. 159–161.



At the national level, DATAR attempts to act on governmental and private economic decisions. DATAR sees itself as a broker pulling strings and getting people together to structure decisions in a more regionally rational way. The state has a vast assortment of selective subsidies and controls it can apply according to regional planning criteria. The job of DATAR is to get the state to do that. Included in this catalog of state powers to "command" by carrot-and-stick procedures are subsidies, fiscal advantages, indemnities, loans, and other devices. Control over building permits is possible, although as a means of weeding out good projects from bad, even DATAR finds it difficult to apply. According to Bernard Pouget only one or two of fifty projects is blocked, and in January 1966 only two unfavorable judgments were made out of 400 applications.<sup>51</sup> In these direct state controls DATAR's role is more often than not that of prodding the Ministry of Finance to follow criteria of regional planning.<sup>52</sup>

### Some Conclusions

Regional planning is the latest innovation in the *dirigiste* or "neocapitalist" political economy which emerged in France in the post-World War II years. In France perhaps more than anywhere else in the advanced capitalist world, the state has come to play a key organizing and energizing role in economic life. Because the very rapid economic development that flowed from these planning efforts tended to accentuate regional imbalances already present in France, regional planning became a necessary addition to the planning operation.

The institutional evolution of regional planning paralleled the evolution of its aims and approaches. It began in the Ministry of Construction with the limited objective of organizing the provision of urban infrastructure to accommodate immigrants and quickly grew to include efforts to promote industrial decentralization in order to redirect that migration. As emphasis shifted to a still broader notion of fitting a geographical dimension to national development, regional planning moved to the Commissariat du plan, an institution that represented a broader perspective and a much less direct approach to getting things done than the ministry. The new objective and approach were finally formalized by the creation of new institutions: DATAR on the national level, and a new administrative structure on the regional level. DATAR attempted to reproduce for regional planning the kind of lightweight, "broker" administrative body which the Commissariat du plan had proven to be effective for national, sectoral planning.

Regional planning emerged, as national planning had before it, totally outside the parliamentary political sphere. It was an administrative operation, aimed at other administrative bodies and private economic interests. Direct participation by elected representatives was minimal. Moreover, French re-

<sup>51</sup>Pouget, p. 92. See Minguet, p. 59, for analysis of building permits issued.

<sup>52</sup>For a summary of these efforts, which helps, although it raises more questions than it answers, see *Budget annexe: rapport sur l'exécution du plan*, vol. III.

gional planning was an attempt at injecting some geographical rationality into national economic planning, not an attempt at giving the grass roots more say about economic objectives. Thus it involved a *centralization* of influence, not the regionalization or decentralization of power as one might naively have expected.

The events of May 1968 and their inconclusive aftermath provide a useful point for terminating our overview of the development of French regional planning. The May events were a huge catharsis for much of French society, a few weeks in which large numbers of Frenchmen expressed their grievances in spectacular ways. And it was apparent to many—including General de Gaulle—that a good many of their grievances were “regional” in nature. The labor conflicts which detonated the wave of sit-down strikes originated and were most militant in areas where regional complaints were strongest and most justified. Labor in the Nantes–St. Nazaire region (where declining shipbuilding had not been replaced by significant new economic activity) was the spark for the whole movement. The Renault plants that played such a key role in the evolution of the strike (Cleon, Flins) were shining examples of industrial decentralization—and of the types of problems that it brought. Moreover, at Nantes and elsewhere the May strikes triggered eruptions of regional grievances that extended far beyond the workplace and the working class. The regional grievances expressed in May were not always narrowly economic: resentment of the destruction of the whole social fabric of provincial France surfaced, along with exasperation at the centralized rule by technocrats who were responsive to their own administrative rationality and to the will of an aging general, but who were not responsive to the people in any way. The problems of regional planning are part of the broader problems of postwar France. The economic expansion, indeed transformation, of France since the war has been a great success story. But it was bought at the expense of severe economic imbalance, social dislocation, and political alienation. May 1968 was a spectacular warning that these problems had to be dealt with.

De Gaulle was quick to realize that something had to be done. In the midst of the crisis he addressed the nation and promised to bring “participation” to France: “participation” by workers in industry and “participation” by citizens in government on a regional level. Early in 1969, de Gaulle submitted a referendum to the country asking Frenchmen to pronounce, yes or no, about the introduction of new regional political institutions. He took the occasion to tie a none-too-popular reform of the Senate, a body he found particularly irksome, to the regional referendum, and he presented the whole package as a plebescite on his regime.

The regional reform package was defeated at the referendum. More precisely, regional reform was not defeated—de Gaulle was. Regional reform went down with him.

But only temporarily. May took the “regional problem” out of the quiet politics of economic administration and made it a public issue. The refer-

endum pushed it further into the center of popular politics. A literature, both technical and popular, suddenly snowballed. Every major political party and group began to take a position on the Regional Question.

The Gaullists were the first. May deprived them of de Gaulle, thereby forcing them to try to establish solid bases of local support and organization. Regional reform provided an excellent occasion to undermine the power of local *notables* and replace it with local Gaullist political organization. So for the Gaullists, regional reform was a way to create at least an illusion of participation while building desperately needed organizational infrastructure for the party.

Jean-Jacques Servan-Schreiber, the new leader of the Radical party (and old editor of *L'express*, France's *Time* magazine), made regional neglect the central issue in his dramatic electoral campaigns in Lorraine and Bordeaux. JJSS, whose initials are a household word in France, seems to have failed in his central aim, to turn the Radical party into the core force of a Center-Left coalition (in fact the Radical party split apart under his leadership), but his commanding use of the media played an important role in making regional questions prominent public issues.

The organized Left took longer to find the regional question interesting. Regional autonomy and particularism, the preservation of local cultures and especially the importance of preserving a healthy, independent rural base against the menaces of urbanization had long been major themes of the extreme Right in France. But they are now being taken over by the Left, which is raising the basic questions that were carefully sidestepped during the administrative phase of regional planning.

France's rapid postwar economic development has created enormous stresses in the regional fabric of French life. Traditional regional structures and identities have been largely destroyed along with the old agricultural France and the old industrial economies. The social remnants of this transformation are politically explosive; so too are the socially threadbare structures of the emergent new provincial life. During the long administrative phase of regional planning, timid technocratic efforts to cope with these problems fell far short of what was necessary. Since May 1968, the regional question has been forced out of the politics of economic administration; it has become an important public issue. Just how the new politics of regional planning will affect the regional structures of French life remains to be seen.

## 35 **Methods of Achieving Consistency between National and Regional Location Plans**

D. Schejbal and O. Žurek

Consistency of the aims and efforts expressed in programmes and plans for the development of a national economy and purposeful regional arrangement are an inseparable part of national economic planning and programming. They are the most substantial part of regional planning. Production location concerned with the branch viewpoints and needs as well as the conditions of individual regions is a complicated and exigent economic activity. The correct execution of this work can benefit the producer and society as a whole. It is necessary to follow all relations induced by a newly located production plant in each region and to evaluate these relations objectively.

The theory and practice of national economic planning are continuously facing new problems and tasks. The following study presents the basic criteria used to approach the solution of such problems in national economic planning in the Czechoslovak Socialist Republic. The main objective of this paper is to contribute to a mutual recognition of efforts for the development of planning methods and management activity in the economic life of various countries with different socio-economic orders.

### **The Relation of the National Economic Plan to the Development of Economy in Regions**

#### **General Conception of the Problem**

The economic development of a country involves problems related to the location of the economy. The degree of solution of location differs considerably in individual countries and in groups of countries. In various socio-economic orders economic laws operate in different ways and their influence on the production process as a whole also varies.

In centrally planned economies the possibilities for the solution of questions related to the location of the economy are favourable because the government controls the planned management of the entire economy and employs it for the benefit and satisfaction of society as a whole. In market economies there is a vast scope of interests and goals which cannot always be harmonized.

The solution to problems related to the achievement of consistency between a national plan and regional location plans depends on the character

of the development of the national economy of a country; that is, whether planned management of national economy is involved or not.

A scientifically planned economy is one that develops according to a programme drawn up in advance, rationally utilizing all available possibilities and means. A scientifically planned economy not only considers all spheres of the production process such as manufacturing, distribution, exchange and consumption, but also all aspects of material, labour, costs, sectorial and territorial planning, as well as the time aspect.

The criteria for a planned economy must then have these features: a. proportional development on the basis of socially conscious planned activity of the people; b. consideration of all spheres of the production process; c. developments for the interest of society as a whole with an aim to raise living standards continuously—both material and cultural.

Regional location of production is an inseparable part of economic planning in the development of a national economy. The position of regional planning in the complex system of the development of a national economy, especially the location of industry—technically and economically the most progressive sector of a national economy—becomes an inseparable part of a national economic plan and one of its most important aspects. The degree of consistency between a national economic plan and regional location plans—especially where industry is concerned—is dependent upon the degree of planned management of the production process.

The necessary starting point is the development of production. Changes of proportions in the national economy as a whole are reflected in the proportions of various groups, branches and industries in one region or another, and in the proportions among individual regions of the country. Correct relationships between branches and industries within each region and the creation of missing links or the fortification of backward links in regions is of extraordinary importance not only for the development of a given region but for the economy of the entire country.

The proper regional location of production means the harmonized development of all branches of national economy in the regions and the creation of correct proportions in the development of the whole economy and culture. Correct proportions contribute greatly to the effective development of the national economy and help to create conditions necessary for the development of higher living standards. To achieve such proportions it is first necessary to consider the possibilities and needs of the national economy and the possibilities of rational exploitation of the natural and economic conditions of each region. This information becomes the basis for regional economic planning, and within its framework regional location plans are formed.

Proportional development of all regions of a country does not mean a simple arithmetical balance, for example, of the rate of growth of industrial production, the structure of industrial production and the level of per capita income. Proportional development of all regions signifies a proper division of labour among individual regions and a purposeful harmonized development

of individual branches in the regions. This is achieved by preferential development of those branches and industries of national economy in regions that have optimal conditions for their development. These conditions can be of a natural or economic character. To ensure the most purposeful growth of national economy, each region must utilize to a maximum degree—but rationally from the point of view of society—all its available resources. The economy of all regions can then develop in a rising curve, even if it is not possible—or even purposeful—to develop at a uniform pace. At the same time it is necessary to continue efforts to eliminate substantial economic differences among regions.

Regional proportionality in the development of a national economy is not only a matter of proportion in the economic development of the regions themselves, but of proportion that will bring about the purposeful exploitation of natural and economic conditions for all regions by the exploitation of national wealth, natural and financial resources, the labour force, and all other resources and interrelationships.

The principal relationship between a national plan and regional location plans is one of basic proportions of national economy in the territorial aspect of individual regions and among the regions mutually. This implies comprehensive plans for the economic development of regions with a mutually harmonized development of industrial and agricultural production, transportation and building, trade and commerce, cultural, educational and sanitary facilities and housing projects and the solution of such problems as water supply.

The location of production, one of the substantial factors of economic development in regions, must be examined comprehensively from the viewpoint of development of production in regions and their territorial relationships. Such an examination will express the conditions and costs connected with the implementation of regional location plans.

These considerations lead to the term “complex economic development of regions,” which is often theoretically overvalued but only partially understood.

Complex economic development of regions can be understood as the proportional development of: a. a complex of branches of a national economy, that is, certain parts of the productive sphere; b. the necessary complex of branches of the non-productive sphere.

The complex economic development of a region has individual links forming the base of the proportional development of its economy. It is possible to analyse these links and to determine the proportions of development of a region and the position of its economy in the national economic system.

Experience has shown that certain general principles are valid for practical application in proper location of production. Economic activity may be subordinated to practical application according to specific conditions in each country. These principles are generally valid, though only one of them may be accentuated in each country at certain stages of development.

The general principles have been developed from the fact that the rational location of production in the regions of a country is reasonable only if favourable conditions are created for: a. the growth of social productivity of labour; b. the rational exploitation of all resources of a country such as natural wealth, labour and financial resources; c. the solution of socio-political problems connected with overcoming economic backwardness and conspicuous differences in living standards. (The backwardness of certain territorial entities must be solved by the government through an effective regional policy.)

These general principles for the location of production express not only the goals of practical activity but also the point of departure for the achievement of consistency between the national economic plan and regional location plans.

#### **The Branch Principle in National Economic Planning and the Ensuring of Regional Proportionality**

Consistency between the national economic plan and regional industry location plans from the viewpoint of the principles and criteria mentioned above requires an organizational structure of the national economy as a whole, and managing and planning authorities. Working together they serve this purpose to the best advantage.

Economic proportionality in the development of a national economy requires: a. a development plan for all economic branches of the national economy, evaluation of the role of sectors, plans for the utilization of labour, and other detailed planning prepared by a state planning commission; b. development plans for individual branches and industries worked out by central organizations (ministries) and their subordinate organizations; c. development plans for the economic activities of regional, district, municipal and local administrations.

The comprehensive development plan of a national economy is thus formed, apart from the basic economic balances, by a combination of comprehensive development plans for all economic branches and industries according to their organizational structure and according to the development plans of the regional branches.

In practice the organizational structure is expressed in the so-called two lines of planning (even if several planning levels are involved): a. planning of economic activity by central organizations (branch planning bodies); b. planning of economic activity by regional organizations (local, municipal, district and regional).

In the system of national economic planning the branch principle prevails and will continue to do so. This principle does not offer a true image of its feasibility because the verification of regional possibilities cannot be realized; for this reason an examination of the regional aspects of the development of individual branches and industries must be made. Consistency of purpose of a national economic plan and regional location plans may be assured by:

- a. Profitable division of labour, to ensure regional proportionality of economic plans on all levels of management and planning; and to ensure close mutual co-operation to determine the duties and liabilities for procuring basic data;
- b. Establishing specific responsibilities for subordinated organizational units such as branch managements, branch enterprises and research and project institutes;
- c. The creation of a system of methods by means of indices, of economic and financial instruments of planning documentation to study economic phenomena, to establish production location and ensure economic and cultural development of the regions in accordance with their specific conditions;
- d. The determination of appropriate economic and financial instruments to ensure regional proportionality.

#### **Economic Territorial Division of the Country**

To ensure advantageous location of industry in accordance with the national economic plan, it is necessary to divide the state into units in which the planned management of national economy from the territorial viewpoint would be possible. These territorial units should correspond to the requirements of the study and the expression of economic and natural phenomena.

Such a division of the country into regions according to economic and natural conditions determines the participation of the regions in the division of labour and creates favourable preconditions for the planned establishment of proportions on the basis of rational exploitation of resources.

Natural and economic conditions in individual parts of the state vary greatly, resulting in substantially different regional units linked by a rational division of labour.

The division of the country into regions can then be carried out as follows:

- a. Economic regions are created according to economic and natural conditions and are used exclusively for the solution of questions related to the implementation of regional plans of production location. These regions usually have no directive organs, and the activity concerning their development is carried out centrally. The regional location plans are then worked out according to these economic regions.
- b. The regional location plans can be based on the administrative division of the country (regions, districts, municipalities, etc.). The advantage of using the administrative division of the state for regional planning lies in the fact that many directive and planning bodies are organized according to the regional aspect of the development of national economy so it is possible to organize co-operation among lower planning bodies. For example, the division of public administration of the CSSR as created in 1960 became the basis for the development of works connected with regional proportionality of national economic plans. This administrative division of the country into ten regions and the capital, Prague, and within these regions 108 districts, is practical for regional planning. In creat-



ing this regional organization attention was paid to the decisive economic conditions of separate regions as well as to the possibilities of effective development of the economic-organizational function of the state in regions.

The division of the state into smaller units making possible the examination of all areas of operation is one of the conditions needed for the achievement of consistency between the national plan and regional location plans. Such a territorial division not only helps to solve economic problems effectively but brings about a comprehensive cultural development of all regions.

The merger of centres into larger territorial units differs according to their problems. It is difficult to solve the balances and distribution of manpower, but such balances must be worked out not only for districts and regions but in relation to principal economic centres and transportation facilities.

To ensure the regional aspect of the development of a national economy, it is useful to examine a whole system of basic data and analyses according to chosen territorial units (regions) which will differ from each other in the extent and intensity of their natural and economic conditions.

The economic region is therefore a general denomination of various types of territorial units. In the Czechoslovak Socialist Republic the public administration division of the state has created the basic regions for which economic plans are drafted. Slovakia represents a territorial unit within the framework of a united Czechoslovak economy in which there are specific economic and political aims and a national development and historical background different from the regions of Bohemia and Moravia.

#### **Consistency between Regional Location Plans and the National Economic Plan**

The achievement of consistency between regional location plans and the national economic plan as an inseparable part of the whole process of planned management of national economy is realized in five basic phases: a. collection of data; b. analysis of data collected; c. decision-making: setting of tasks, determination of instruments and conditions; d. fulfilment of tasks under operational management; e. final control.

##### **Collection of Data**

The beginning of the process of planned management is the collection of such data as the condition of the national economy, the level of productive forces, the state of exploitation of natural resources of the country and the influence of national and international political factors on the development of the national economy. The collection of such data from the regional aspect has a decisive influence on the determination of a starting basis, and for the estimation of basic trends and tendencies of development in regional location.

##### **Analysis of Data**

The analysis of this regional data is the second phase of the process of management. Careful analyses contribute to: a. qualitative examination of individual phenomena and their individual components in regions; b. determination of the decisive links and activities, and interrelations among indi-

vidual phenomena in the regions; c. formulation of conclusions about factors influencing the development of the phenomena examined, and measures to be taken in accordance with the aim of the analysis.

From the regional aspect, thorough scientific analyses of the state of location and of the natural and economic conditions of individual regions are fundamental tasks at this phase of management.

These analyses should be elaborated at the beginning of the work on a long-term prognosis of the development of a national economy. They are of a twofold character and are worked out along two lines—branch and regional.

Central branch organs work out, with the help of their research and project institutes, the analyses of present development, the state of production location in individual industries, development tendencies, the scientific and technical developments of industries and their influence on the further development of production location.

The regional organs work out summary analyses of the achieved degree of development and the exploitation of natural and economic conditions of individual regions.

On the basis of these analyses and its own data, information and studies, the planning centre determines: a. the possibilities of further exploitation of natural and economic conditions of individual regions; b. the factors which can accelerate, decelerate or even limit the development of production in the regions; c. the main changes in production location which may be expected in the coming years as a consequence of present or already started development and as a consequence of the development of new techniques or other development tendencies or targets; d. the problems to be solved.

The basic data for these analyses can be divided into such categories as: data about natural conditions, raw materials and emergency resources (the possibilities of their exploitation and the probable time when exploitation can begin); data about demographic development and structure of the population; the location of industrial and agricultural production; production and consumption relations within each region and among regions; and transportation data.

The data should be suitable for perspective plans, the material clear, of simple construction and with graphic and cartographic annexes. This basic material must be available to the decisive branch and regional planning bodies.

The development and the predetermination of industrial location in the Czechoslovak Socialist Republic and the basic tendencies of influence must be respected to a large extent in further plans of regional location of production. That is the case of the Czechoslovak Socialist Republic at present, where a long-term projection for fifteen to twenty years is being prepared. The conclusions arrived at, on the basis of the analyses, are applicable to a number of industrially developed countries; they are:

- a. The present development of Czechoslovak economy demands exploitation of conditions and resources and the utilization of already existing

productive fixed assets for the achievement of a more regular location of production, so that the future development of national economy and, with it, production location will be influenced by the degree of exploitation of these conditions for economic development. Economic relations within the regions and among them have increased, becoming significantly more intricate.

- b. The necessity to increase economic effectiveness through intensification and through the realization of necessary structural changes, especially in manufacturing industries, tends to lead to a situation under which the increase of production will be further concentrated in key plants, branches and industries.
- c. The location of new productive capacities in a number of branches, especially in manufacturing industries, has to be secured—while observing the principles of social effectiveness—in a manner that utilizes manpower resources available in certain regions.
- d. The necessary process of reconstruction and modernization of productive fixed assets of a number of branches will have an outstanding regional character. The need for complex development of the regions will be an important condition in decisions on the order and significance of reconstruction and modernization of existing productive fixed assets (the problem of the reconstruction of productive fixed assets in Bohemian regions in districts and branches industrialized in the epoch of the primary industrial revolution).
- e. The process of production concentration and specialization already begun in many industries (engineering, textile and garments, and food-stuffs) brings new regional problems in the full exploitation of new productive fixed assets, in the growth of new and more complicated inter-industry relations and in the problems of reserve labour opportunities in localities where old productive capacities have been liquidated.
- f. Development of agriculture requires not only selection of land suitable for intensive development in individual regions, but the solution of a whole set of problems of outstanding regional character brought to life by the introduction of wholesale production in this sector.
- g. The organization of cargo transport in national economy as a whole will require a strict observation of regional aspects in the location of productive forces.
- h. The direction of the flow of supplies resulting from the developed international socialistic division of labour (coal, iron ore, oil, gas) became a new factor in the rational and effective location regions (Bratislava, Košice) where extensive construction of new productive capacities of a series of basic industries is under way (metalworking, chemistry and engineering).
- i. The development of production occurs under continuously deteriorating conditions as far as water supply for industry, agriculture and population is concerned. The maintenance of water in a region requires good care of forests and the avoidance of pollution.

- j. It is necessary, more now than ever, to link the gradual solution of housing problems with the elimination of backwardness in technical and civic undertakings.

The preceding survey, which should serve for basic orientation only, shows that these and other circumstances have their influence on the development of the economy as a whole and urgently require a sensitive solution to regional distribution and location.

#### **Decision-making: Setting of Tasks, Determination of Instruments and Conditions**

The third principal phase of planned management of national economy and its regional aspect of development is decision-making: setting of tasks, determination of instruments and conditions for their implementation.

Economic decision-making depends upon the analyses and studies carried out, the goals of economic development, the determination of direction and pace of this development, the basic proportions of the development and the determination of means for its realization. The basic instrument of management and thus of decision-making is the national economic plan as a general conception of economic development, giving the proper orientation to each enterprise as a part of the national economy. The plan must represent the binding framework of the economic policy of the state, specially through centrally directed financial credit, income and price policies. The national economic plan and measures taken by the state in this respect will determine the concrete tasks of the units of economic activity.

Decision-making should be considered the main phase of the process of management in the field of regional planning, and the choice of indicators and instruments should be subject to it.

#### **Fulfilment of Tasks under Operational Management**

The fulfilment of the plan and the realization of economic tasks occur under ever changing conditions. In the fulfilment of the plan itself unforeseen circumstances, unsatisfactory work, the influence of foreign trade, and changes in international relations, can alter the economic relations of future development and, consequently, influence investment activity and the development of the labour supply.

It is therefore necessary to adjust or to modify the set economic tasks according to influences, and to do so on the basis of influences on operational management of economic administration.

#### **Final Control**

The control of the fulfilment of the tasks by the managing organs which closes the process of management is important. While controlling the fulfilment of the development plan of national economy from the regional viewpoint, it is necessary to concentrate on the principle of the main link, to direct attention to the main link of the national economic plan, to regions and tasks which can influence in a decisive manner the goals and development of the national economy.

For this purpose it is necessary to establish a system of indicators and

controls. Individual changes occurring in the regional production location usually have a long cycle; it is therefore not necessary to execute the control of the fulfilment of the plan in this field in short-term intervals: semiannual and annual controls and usually sufficient. In some economic activities the time period can be longer.

Control should be based on the data of statistical services, which should organize their work in relation to regional problems.

### **Organizational Forms of Planned Management**

The solution of problems related to the realization of economic plans from the regional viewpoint demands effective organization.

In the system of planned management of national economy, the basic condition for the solution of the regional aspect of the development of national economy is the creation of relations of the planning centre with central branch and industrial organs and regional organs.<sup>1</sup> Thus, it is necessary to create the conditions for the elaboration of the draft of the plan in individual branches of the national economy (the co-operation of central branch and industrial organs and enterprises with regional organs). These relations create the organizational conditions for the estimation of proportions in the development of economy managed by the centre and economies managed by regional units.

The basic proportions in the economic development of regions, the proportions among regions and the use of economic instruments must be determined by central authorities. The regional aspect of the development of a national economy is, therefore, directed and centrally co-ordinated by the planning centre (in the CSSR the State Planning Commission). For this purpose it elaborates the analysis and evaluation of the existing state of exploitation of natural and economic conditions of individual regions, and works out basic proposals for effective exploitation.

On the basis of these analyses and considered lines of development of the national economy, it proposes basic long-term directions and aims for the development of individual regions consistent with the needs of further development of productive and non-productive spheres, and presents these proposals to the Government for approval as part of a comprehensive draft of the national plan. Ministries and other central units, branch managements and association enterprises, and regions are informed of these directions and aims. The proposals of the centre are then compared with the technical-economic studies worked out for the development of individual branches from the viewpoint of economic needs and requirements. On the basis of these comparisons, basic long-term directives and aims are set.

<sup>1</sup>In the CSSR the law emphasizes the highly responsible role of national committees—organs of popular administration—in securing the harmonious complex development of national economy in regions administered by them, irrespective of the subordination of economic activities in the regions. In this sense the term “regional organs” is used in the text.

The planning centre has a special department of regional proportionality which elaborates and applies appropriate economic and financial instruments. These form part of the whole economic system which ensures regional proportionality.

The planning centre then proposes a general system for: co-ordinating the participation of ministries, central units, branch managements and regional units for material and organizational problems of production location and economic development; the co-operation of central statistical units to determine the required accounting and statistics and to present suggestions for scientific research concerning these problems.

Central and branch units work out calculations and analyses for development and changes in location of the economy with which they are concerned from the viewpoint of the location of production, fixed assets and basic co-operative relationships.

The development of branch industries in regions usually is determined by:

- a. The conception of location on a productive-technical and development-research basis based on existing location and maximum exploitation of existing capacities;
- b. Classification of organizational units, enterprises, plants, works in regions, that is, the division of individual enterprises, plants and works into categories of development units, stabilized units, units for liquidation, expected changes in the utilization of productive capacities, and the like;
- c. The conception of the organization of the production of the branch in the region, that is, especially the problems of concentration, specialization and combination of productions with respect to customer-supplier relations and co-operative relations and the conception of the development of technical level of production;
- d. The conception of the development of the production of individual regional branches and industries on the basis of economic needs;
- e. The utilization of economic instruments for ensuring long-term development of production of individual branches and industries consistent with regional proportionality.

Regional units are the necessary organizational components to solve regional problems in the development of the national economy, and they should ensure the full economic development of the regions.

In accordance with the needs of the national economy on the basis of instructions from the planning centre and the Government, regional units can ensure the most purposeful use of natural and economic regional resources and the consistent development and location of the economy, administered and planned by local organs, with the location and development of the centrally managed economy. They must estimate the development of proportions and relations in the complex economic development of regions, smaller territorial units, industrial agglomerations and large cities.

Regional units must work out comprehensive draft plans for economic development of the regions on the basis of: directives from the planning

centre, their own basic data, data from branch and industrial managing organs and their subordinated units, from communities, municipalities and districts for enterprises administered by them, and present these drafts to the planning centre. A comprehensive draft then becomes a basic document for the management of economic activity.

The suggestions of regional units for better utilization of their natural and economic resources, for achieving interregional proportions, and for utilizing economic instruments become a part of the draft of the over-all plan.

### **The Regional Aspect of National Economy and Planning Documentation**

The system of national economic indicators, economic instruments, planning methods and procedures and documentation help to develop the regional aspect of a national economy. To fulfil its mission the system must respect individual conditions of branches, industries and regions so that specific features are observed and unified activity achieved.

#### **System of Indicators**

Economic indicators in the field of regional planning are the verbal expression and numerical determination of the quantitative contents and qualitative relations of one aspect of the production process—the regional aspect. This system should:

- a. Determine the basic relations and the degree of utilization of natural and economic conditions of the regions;
- b. Estimate and determine the main changes in production location and the basic proportions of the economic development of regions, mainly the proportions in the development of productive and nonproductive spheres, between the development of industry and agriculture, between the need for construction work and the capacity of building industries, between the resources and need of manpower, and between incomes and expenditures of the population;
- c. Estimate and determine interregional proportions in the development of production, transportation and housing;
- d. Carry out, balance and unify work needed to ensure proper regional proportions of manpower, investment construction, water supply administration, transport and incomes and expenditures of the population;
- e. Estimate the economic effectiveness of proposed solutions;
- f. Create conditions for organization at the branch and regional management levels to collect data on responsibilities.

The system of indicators can be divided into two basic groups: a. indicators that become the subject of co-ordination in individual regions such as the indicators of the plan of labour, investment activity, construction work, transport, water supply administration, housing projects, educational, cultural and recreational facilities; b. indicators that estimate and determine the pace of development of the economy in regions, the development of living standards, the state and degree of the utilization of natural wealth and the

state and development of the location of production branches and industries. Together with the first group, these indicators estimate basic regional proportions. Both groups of regional indicators are an inseparable part of the method of instructions issued by the planning centre for drawing up plans for regional and branch organizations.

#### **Economic Instruments**

Economic instruments are also an important factor in the system of methods. Following the basic principle, economic and financial instruments do not work against the major requirements of society in production location.

The basic instrument of the management of the regional aspect of the development of a national economy is the plan with its basic goals, ways and means; however, another instrument of management can be the purposeful connexion of the plan with the market mechanism.

Economic instruments must be used either as a stimulant or a limitation. The choice of these instruments must respect the fact that economic processes in regions and localities demand the expenditure of national resources (induced investment).

The basis for the determination of economic instruments to ensure the regional aspect of the development of a national economy is the analysis of the development of the economy in regions. It is determined by a long-term (five-year) plan indicating problem areas and basic goals.

Economic (financial) instruments can be divided into positively stimulating economic instruments, and limiting economic instruments. Some examples are given below.

#### **Positively stimulating economic instruments.**

- a. Subsidies from the Government for partial coverage of investment costs of the investors. The subsidies can be a fixed per cent of the budget cost of the investment (e.g., 15 per cent of the total budget cost) or a per cent limit (e.g., 5 to 25 per cent of budget investment cost when giving preference to certain industries), that is, for regions and localities determined by the Government for a longer planning period, such as a five-year period.
- b. Reduction in the per cent of contributions (taxes) payable from the productive and non-productive fixed assets to level out the increased operational costs of the enterprises in special cases and regions determined by the Government. The enterprises and plants would pay the contribution (tax) from fixed assets in specified regions or specified branches according to a reduced percentage rate (reduction by one third) and for a certain period after being put into operation (3 to 5 years) or for the period of the necessary trial-run of the production.
- c. Credit preference and eventually a reduced interest rate on the credit granted—a certain percentage cut determined in advance for investment operations in selected regions, industries and activities.

**Limiting economic instruments.** Limiting economic instruments influence organizations which plan the development of production and investment acti-



vity in regions with intricate technical and territorial problems, especially manufacturing industries that do not require a location related to a raw material or some specific natural condition.

- a. The increase of the rates of contribution (taxes) from the fixed assets of newly constructed production capacities for a certain time after the start of production (one third of the rate). The resources gained in this manner can be used for the administration in districts, municipalities and localities or for subsidies for the promotion of economically less developed regions.
- b. The determination of additional contributions (taxes) from gross income or profit for a certain period after the start of the operation of production organizations which can be passed to the state budget or to local administration budgets.
- c. The determination of fees in the case of new constructions, where the basis would be a percentage from the price of the realized investment to the state budget or to the local administration budget.
- d. The introduction of an economic instrument to direct the development and proportions in regional distribution and utilization of manpower resources. The utilization of this instrument can be bound to the extent of the total of wages paid out by enterprises and plants (e.g., 2 per cent) in relation to the increment of workers in comparison with the preceding period (that is, a year). This additional fee (tax) can be determined only for regions and localities where there are large differences between the resources and demands for manpower, eventually differentiated to ensure a purposeful development of the structure of labour in relation to resources (men-women) and where the society does not have any interest in further concentration of production and population into a few industrial agglomerations, but wants to solve further location of industry from the political-social aspects. The aim of the application of this instrument is to lead the producers of the given region to use progressive technical solutions in production to save labour and thus to ensure the planned result in production increase.

The use of economic instruments for ensuring the regional goals of the development of a national economy is an important link in the achievement of consistency between the national economic plan and the regional industrial production location plans.

#### **Kinds of Plans**

A long-term perspective plan for the development of a country covering a period of ten or even fifteen to twenty years can deal fully with the problems of production location. Based on the evaluation of natural conditions, the perspective plan can call for: the solution of complex economic problems in regions, the preparation of large territories for industry or agriculture, the exploitation of natural resources such as mineral wealth, the orientation of agricultural production, and the creation of preconditions for industrial

processing. The long-term plan can assess in the best manner the need for and the consistency of activity in the non-productive sphere, the conception of housing construction and the construction of civic and technical facilities. The regional location of production (enterprises or plants) depending on the state of production location in individual regions is preceded by basic studies from the planning centre.

In preliminary technical-economic studies for the development of branches, it is possible to assess the effectiveness, needs and demands of the branches to be located in regions. The determination of technology for the development of industries is one of the basic factors connected with regional problems.

Perspective plans worked out, for example, for a five-year period must gradually bring to a realization the production locations contained in a long-term plan. Five-year plans include a more detailed consideration of regions than is possible in a long-term plan. They estimate the demands for production in relation to space and time and include a selection of location variants. They estimate in detail the relationship between primary industries and manufacturing industries.

Perspective plans are worked out for each year of a given period in a manner which enables mutual co-ordination of time and planned action in regions.

In annual plans the questions of regional development and production location are only mentioned. One-year plans are important instruments for the realization of regional intentions and for disclosing existing or future partial disproportions in regions which have not yet been recognized in detail.

The system of national economic plans used at present in the Czechoslovak Socialist Republic provides every opportunity for achieving consistency between the interests of the economy as a whole and the interests of regional production location. The tasks of individual plans of this system are expressed in their specific methodology, so that the solution of regional problems may be worked out gradually and concretely.

#### **Mutual Co-operation of Planning Bodies**

If a plan is to realize its objectives for the development of the economy, the managing units must co-operate at all stages of the national economic plan (irrespective of its kind) in a time continuity that enables all units to fulfil their functions.

After the determination and issuance of directives for the drawing up of a plan by the planning centre (Government), or during the period of preparation of perspective conceptions, the branch units (ministries, association enterprises and branch managements) must present their projects within the determined scope of regional indicators both to the planning centre and to regional units.

The regional units should participate directly in the preparation of the plan. On the basis of their own data and the data of the branches, they work out the drafts for the development of the economy and culture in regions and

present these drafts to the regional units of the planning centre. The results of negotiations between branch and regional organs are an inseparable part of the planning documentation.

The planning centre then has at its disposal the drafts of the plan of the branch units, with certain basic data about the regional location of production, and the regional complex drafts of the development of the whole economy in the territory of the regions, including the results of negotiations between these units. The planning centre should be informed of the details of contradictions between units. It is at this moment in the process of the creation of a national economic plan that the closest consistency with the regional aspects of realization is reached from the methodological point of view.

On the basis of the data mentioned and its own material, the planning centre should discuss the contradiction of aims between branch and regional units, work out the final draft of the plan of regional location of the economy as an inseparable part of the national economic plan, and determine the basic goals and tasks in this field.

The comprehensive national economic plan (including its regional aspect) must be approved by the Government. In the Czechoslovak Socialist Republic the Government and the National Assembly must approve the basic goals and tasks for the development of the national economy to become a law.

#### **The Regional Aspect of Development**

The method of the main link is a basic method used in the field of regional problems. Usually, the whole complex of economy in regions, as well as the specialization and position of regions in the national economy, are examined by this method.

The decisive factor in complex economic development of regions is represented by the fields of production which establish the position of a region and its specialization. These are the specializations which will bring about the most rapid development of the productive forces of the entire state from the most effective utilization of local conditions, material and manpower.

The degree of specialization of production can be expressed by the share of the total production of all branches in the regions, by the share in state production of respective branches, and by the share of production in the coverage of the needs of other regions of the country. The extent and direction of the relations among individual regions is determined by the concentration of production and the peculiarities of structure.

The level of the development of productive forces in individual regions is, first, dependent on the scope and pace of development of specialized industries. The economic complex of a region is formed around one or several industries. The complex of industries has to be examined, therefore, in relation to the leading industries.

Recent experience in Czechoslovakia shows that in a number of regions it is not possible to realize an all-round development of industrial production. Because of economic reasons, some industries, enterprises or plants receive

"preference". This means that a longer-term process of more distinctive specialization, in accordance with the needs of the national economy, begins to appear.

For the future, the problem rests in determining the main lines of regional specialization, the main links of development, and in the subordination of all other goals to the development of the regional economic structure in these decisive directions.

The development level of production fields and future possibilities consistent with the uninterrupted development of the specialized regional production are taken into account when using the method of the main link.

To increase the level and possibilities of these fields is not only important to mobilize the resources in each region, but also to accelerate the production process, to decrease production costs and to create effective management of production.

Industries and enterprises connected with the production of goods for local consumption are an important consideration, especially the production of construction materials, the consumer and foodstuffs industries and agricultural production. Proportional unity of these three groups is important to ensure the complex development of the economy of regions and the correct regional location of production in general.

There is a certain dependency between technical progress and regional location of productive forces which cannot be omitted in using the method of the main link. Technical progress alters the criteria for production location, liberates production from local borders and enables a more regular location of industry and agriculture throughout the country. The achievements of technical progress can be applied best under the conditions of rational production location, that is, under the most purposeful utilization of natural and economic conditions of individual regions.

The influences of technical progress and regional location of production can be briefly characterized as follows:

- a. Improvement of technology of production changes the demands on production location and the criteria for location.
- b. Mechanization and automation leads to a considerable increase in productivity of labour and growth of production and a decrease in production costs.
- c. Development of techniques enlarges and enriches the raw material basis of an industry; new sorts of raw materials appear; and research helps to discover new natural resources.
- d. Concentration of production increases the specialization of the regions and their share in total production.
- e. Increase of the efficiency of technological methods has a direct influence on the growth and proportionality of the economic development of the region.
- f. Changes in structure occur, such as combined enterprises which completely utilize raw materials and waste materials.

- g. Economies in raw materials made possible by general technical progress of the industry cause changes in the needs of regions.
- h. Changes in transport utilities either heighten the efficiency or change the mode of transportation as an important factor of rational transportation.

A final group in the complex economic development of regions is formed by those branches of national economy which create the conditions for the activity of the mentioned groups of production (transport, water supply, energy and the like).

### **Investment Activity**

#### **Preparation and Creation of Investment Plan**

Investment activity, especially the construction of new productive capacities, has a decisive influence on changes in production location and on the pace and structure of the development of production in regions. The reconstruction of existing capacities, their modernization and completion, have a significant regional importance. The plan of investment construction has become the most important instrument in the regional aspect for the development of a national economy.

The preparation of investment construction can be divided into several time-limited stages according to their character. The initial stage has the character of planning preparation, while the following stages assume gradually the character of project preparation, that is, technical-economic and technical projects and territorial-technical basic data.

The construction of new industrial capacities (enterprises and plants) comes from the fulfilment of development plans of the national economy. Long-term perspectives can solve the complicated questions related to production location, preparation of new sites for the construction of industrial bases, the construction of hydroelectric power stations, railway networks, and the preparation of the territory for the development of agriculture.

By their character, regional schemes emerge from national economic plans and make precise demands on time, material and total investment. Therefore, the result of such a complicated activity is the project for a certain territory which is then decisive in the construction and reconstruction of productive capacities, as well as of the other related objects either induced or conditionally bound to such a construction.

According to the character of the territory and the character of the problems to be solved by the projects, solutions can be brought about by regional schemes for large territories or development plans for settlements, towns, agglomerations or villages.

Territorial projects serve as a basis for the issuance of building permits for the realization of an investment in the territory. The issuance of a building permit concludes basically the site of location. The unit issuing the building permit (usually district, municipal or local administration) controls, before its issuance, the conditions that determine the stages of the preparation of the investment.

Regional schemes are advantageous for the construction of large areas which call for vast and complicated changes in the surroundings, reconstruction of territories, and the construction and reconstruction of cities.

#### **Technical-Economic Motivation**

An important part of the planning preparation of investment construction is an elaboration of technical-economic motivation of the investment to estimate and determine the main economic and technical indicators.

Technical-economic motivation is usually a part of long-term perspective plans of investment construction and is the basis for working out project and budget documentation.

The technical-economic motivation of industrial construction usually consists of such information as calculations to determine the production capacity or the structure of a production programme, the economic motivation of raw material, the fuel and energy basis, calculations on the relation of the production of the enterprise to the regions of consumption, water supply data, transport information and manpower requirements.

The economic motivation of the site of construction in regional variants requires a separate analysis of investment for the construction of the production capacity and the inducement to invest.

The construction of industrial capacities, especially with large demands on labour, calls for considerable costs for the development of the municipality; such costs influence significantly the total economy of the investment location. Under these conditions the investor has to cover, in substance, the investment required for the development of the municipality. He must also examine the possibility of location in economically less developed regions with manpower resources. For economically less developed regions, the planning centre must carry out the following:

- a. Make a selection of economically less developed regions, evaluate their possibilities and determine a programme for the preparation of these territories for economic development;
- b. Exercise "pressure" for the organization of production programmes in industrial agglomerations for the best possible conditions to stress production specialization in these regions;
- c. Contribute to the organization of aid to industrial agglomerations for economically less exploited regions where industrial development is planned. (Training of experts on the spot for the realization of construction and taking over newly constructed production capacities.)

The complex evaluation of all conditions of the regional location of production capacities and their location in specific territories is a vast activity involving not only the quantification of branch conditions, but the evaluation of individual regional, territorial solutions; correct decision-making; and utilization of the time factor. Each final solution in regional territorial location of new plants requires the analysis of a number of factors which help to create conditions for effective production and an improvement of the social and economic life.

### **Association of Investment in Regions**

Association of investments refers to investors joined by a common interest in the development of certain utilities. This has considerable significance in the construction of infrastructure in a territory. It is to the benefit of the national economy to issue a binding rule which solves in detail the procedure of association of investments, including financing.

Association of investments in one common investment brings considerable savings to all participants, especially in costs connected with operation. A well-chosen participation of regional units of administration results in savings, and it is in their material interest to ensure the whole process of association according to the following principles:

- a. Association of investments should be carried out in regions by economic organizations administered by the branches and by the administrative units or enterprises of production co-operatives, regardless of whether the investment is of a productive or non-productive character.
- b. The association should be voluntary, modified by economic pressure in cases where the effectiveness of an associated investment is doubtful (that is, in the form of refusal of investment credit or its limitation).
- c. The position of regional organs of people's administration is strengthened at all levels by the association of investments, especially their active role and initiative in proposing association, in the approval of the main investor, in determination of the participation of partial investors, not only for the construction but for the utilization of associated investment as well, in solving difficulties in financing and realization of associated investments, and in the division of savings achieved.
- d. It is purposeful to determine the participation of the units of people's administration on the use of savings up to the level of 50 per cent of total savings which go to the reserve fund for the development of the economy of these units.
- e. An active role by the state bank is important in decision-making for an associated investment.

### **Contemporary Tendencies in the Improvement of Regional Planning in the Czechoslovak Socialist Republic**

The foregoing explanation of the methods of achieving consistency between a national economic plan and regional production location is based on the practices and experience of the Czechoslovak Socialist Republic. During the period of development of planning a rational economy in the CSSR extensive attention was given to the development of regions and to planning methods which would achieve such development.

Economic development in the last fifteen years has resulted in more than a fourfold increase in industrial production. At all stages of this development considerable emphasis was given to the role of industrial agglomerations; a large part of the effort for the total growth of economy concerned such centres.

This economic development brought about the construction and growth of production in economically less developed regions, and gave birth to new industrial centres so that no economically less developed territories remain. The preparation of a long-term perspective (up to 1980) presently under way will increase efforts for the solution of regional proportionality.

The Government has increased the role of national committees at all levels—regional, district, municipal and local—to ensure complex economic development in the territory administered by each organization. The fulfilment of this task will ensure a harmonious development of the economy of the territory administered by national committees, whether the production units are managed centrally or by local units (Law on National Committees No. 65/1966). The modifications of the National Committees Law cover to a large extent tasks entrusted to the national committees as organs of the people's administration, especially with regard to the necessity to give greater attention to living conditions in respective regions.

The intricate conditions governing the location of productive forces in the CSSR have called for scientific examination and research, especially from the point of view of perspective planning. These requirements have led to a new evaluation of the role of scientific organizations. The Czechoslovak Academy of Sciences and the Economic Institute of the Slovak Academy of Sciences are examining the theoretical problems of the location of production forces. The extent of regional problems in the CSSR has recently led to the founding of a special research institute for regional planning. This research centre will analyse and evaluate scientific methods for the development of the Czechoslovak economy in regions and will examine possibilities in the regions to determine their role in the economy of the country as a whole.

### Conclusion

This paper has endeavoured to present an explanation of basic criteria, preconditions, methods and organization of regional planning as an inseparable part of national economic planning. The growing exigency for a solution of production location and a harmonious development of the economy in regions consistent with their natural and economic conditions has been pointed out.

The solution of the problem requires a counterbalance between the application of the branch principle in the planning and management of the economy and the ever growing complexities of the methods of regional planning. The preparation of draft plans and their realization require organizational conditions. In this field the problems of management should be connected with economic regionalization of the country. The CSSR uses for this purpose the administrative divisions of the state.

All activity related to the regional aspect of a national economic plan must be based on detailed facts about the regions and a recognition of the demands of branches and industries on the "absorption" of existing possibilities in regions.



Methods for this work are based on the system of indicators, on economic instruments—limiting the development in certain regions—and on the application of measures to regulate regional economic development.

For the proper functioning of the whole system of regional planning, it is necessary to ensure that branch managing organs deal with the regional aspect of the development of their production. It is necessary to authorize the appropriate units in individual regions to organize activity to ensure all-round co-ordination, especially concerning the intentions of various participants in the development of economy in regions.

Technical-economic motivation of the location of productive investment, all-round preparation and implementation of investment programmes, and the evaluation of variants of location of new production in regions aid in the exploitation of investment activity as an important factor in production location.

#### Bibliography

Act on National Committees, 1960.

O. Žurek: *Location of Productive Forces in CSSR*—Státní nakladatelství technické literatury, Prague, 1960.

Statutes of the State Planning Commission, 1963.

Decree on planned management No. 90/1965.

Government Act on the Heightening of Effectiveness in Ensuring Regional Proportionality in National Economic Plans, 1965.

Twenty years of the development of Czechoslovak Socialist Republic—Nakladatelství politické literatury, Prague, 1965.

Planned Economy, 1965/No. 6—D. Schejbal: "Regional Problems in the Project of Development till 1970".

Planned Economy, 1965/Nos. 7, 8—O. Žurek: "Regional Proportionality in National Economic Plans".

Planned Economy 1965/Nos. 9, 10, 11—M. Souček—J. Tauchman—Z. Vergner: "Preparations of Long-term Prognosis of Economic Development of CSSR".

Planned Economy, 1966/Nos. 7, 10, 12—L. Říha—J. Bedrunka—F. Šalda: "Scientific-Technical Development and Economic Structure".

Act on Five-Year Plan 1966–1970, No. 83.

Statistical Yearbook of CSSR 1966, official statistics, planning documentation.

## 36 Aspects of Regional Planning and Theory in the United States

William Alonso

This paper deals with some aspects of national regional planning in the United States. I shall, of course, deal primarily with issues which are explicitly regional, such as distressed areas and urbanization. But, although the popular image is that the United States does not engage in national planning, I shall try to show that territorial planning is intrinsic in the very process of American government. The last pages will deal with recent theoretical and technical developments and their interplay with the realities of application. These issues are fluid, complex, full of life and energy and therefore ambiguous, contradictory, and often surprising. No brief portait can do justice to such a subject, and therefore this survey of the situation must be, of necessity, an impressionistic personal synthesis.

The United States is a country still engaged vigorously in its own economic development. Since 1950, over twenty million people have left rural areas and small towns for the city. Perhaps to a greater degree than any other developed country, its economy consists increasingly of goods and services which had not existed a few years ago. The permanent structural transformation of economic development brings with it sectoral and spatial dislocations which often result in lags or imbalances in the integration of the space-economy. Vast portions of the nation are officially classified as economically distressed by the Federal Government. The areas include agricultural districts depressed by changes in the structure of the market for their products and districts where advances in agricultural technology have outmoded the family farm. They also include towns and cities whose mining base has played out, or whose industry has moved out or become obsolete. The syndrome of such problems is well known. Rural areas, with thinning population, are forced to close some of their schools to consolidate the education of their children in fewer places. Market centers with insufficient demand decline or disappear. Low incomes and high unemployment become typical. Very often there are also serious socio-psychological problems, such as a rural anomie characterized by feelings of diminished self-esteem, apathy and the inevitability of decline.

This paper is based on work done under a grant from the Economic Development Administration of the U. S. Department of Commerce.  
Reprinted from *The New Atlantis*, vol. 1, no. 1 (summer 1969), pp. 8-22, with permission of the publisher.

Often the structure of the family is weakened by the departure of the husband or the children, or because the wives go to work in the low-wage industries which often move into such areas while the husbands find themselves unemployed and dependent. This is not the place to enter a detailed classification of these areas or description of their problems. Rather, I will discuss the efforts to meet them.

The United States has a strong tradition of local self-determination and self-help. A depressed or backward area will typically try to attract industry. A local development corporation is often formed, some vacant land is designated as industrial park, and capitalization may be sought with which to offer subsidies to industry through development of the industrial land or by attractive loans. Sometimes industry is offered reduced local taxes, or the use of the credit of the local government, which, because of certain features of federal and state tax laws, provides a very low rate of interest. Other efforts may include advertising in business publications, soliciting of prospects if likely candidates can be identified, improvements of the attractiveness of the town, friendliness and human warmth, and assurances about the docility and diligence of the local labor force. Occasionally these efforts are successful, and these successes are glowingly reported by federal agencies, by publications such as *The Wall Street Journal*, and by national associations for local development. It is seen as a demonstration that, in America, a community or an individual can achieve success through perseverance and hard work. However, Gilmore, who estimated that there were some 14,000 such public and private organizations in 1958<sup>1</sup>, concludes that by and large they are not very effective. The difficulty, of course, is that local problems are not the result of random bad luck, but tend to reflect changes in the organization of the national structure. Since the small sub-system is largely dependent on the flows and forces of the larger system, local efforts are unlikely to bring about large results, accomplishing perhaps some minor spatial shifting of industry.

Federal programs for aid to distressed regions are similar to those of other countries, with some American characteristics. They rarely designate specific areas. Rather, they define distressed conditions under which areas can qualify for aid. These conditions are low income, high unemployment, persistent outmigration, or a combination of these three<sup>2</sup>. Qualifying areas are entitled to participate on their own initiative in a wide variety of federal programs which provide funds for planning, loans or grants for construction of public utilities, roads, schools, health facilities and other forms of infrastructure, preferred treatment on business loans, and, increasingly, worker-training programs and aids to management.

The limitation of resources available for such aid requires the

<sup>1</sup> D. R. GILMORE, *Developing the « Little » Economies*, New York: Committee for Economic Development, Supplementary Paper No. 10, 1960.

<sup>2</sup> I omit here other less general criteria such as natural disasters, or a very high proportion of school children from federal employees.

establishment of administrative priorities in meeting local requests. It must somehow be determined, usually by informal and general understandings at the federal level, whether it is more urgent to deal with areas of low income or with areas of high unemployment, or whether the special problems of Indians or Eskimos deserve priority. Recent United States practice has been on a basis of « worst first ». That is to say, after the qualifying areas have been ranked by some criteria, however arbitrary, preference is given to those which appear to be worse off.

The implicit logic of this practice bears a moment's examination. Although there are subsidiary goals, American regional policy, as that of most other developed countries, addresses itself to two principal goals. The first is that of efficiency or national economic growth. The second is that of equity, or a more equal distribution of incomes, jobs, and other benefits. The language of federal legislation implies that existing programs satisfy both goals, yet it is obvious that alternative policies will favor one or another goal to different degrees. For instance, it seems extremely likely that in general the marginal return to investment in depressed areas is lower than in prosperous and dynamic ones. Therefore, legislation which encourages the investment of national resources in depressed areas would seem to sacrifice some national economic growth for greater internal equality. Within the pursuit of equality, the « worst first » policy seems to assume that the marginal utility of income in the poorer of two distressed areas is sufficiently greater to offset its greater marginal capital-output ratio. It is clear that it is impossible to determine whether this implicit assumption is correct<sup>3</sup>. Even if we knew the marginal capital-output ratios, which we do not, how can marginal satisfactions be compared? Theory and fact can be advanced further than I have done in this illustration, but the point is that technical analysis is a limited tool to evaluate particular equity strategies such as the « worst first »<sup>4</sup>.

Most of the American legislation addressed explicitly to territorial issues may be said to be problem-oriented, like medication for a localized disease, rather than oriented to guiding national development. Economic difficulties in a district are viewed as a pathological condition rather than as national growing pains or stresses of adjustment to a permanent process of growth and change. But there is a hidden portion of the iceberg of regional policies in programs whose legislation carries no direct reference to regional issues, and many of these programs address themselves to developmental questions. As in most democracies, the executive branch of the American government is organized into ministries (Departments)

<sup>3</sup> This is my own interpretation of the assumption. An alternative assumption would be that marginal capital-output ratios in fact decline with economic distress, in which case a « worst first » policy would serve the efficiency goal as well. However, if this were the case, it is hard to see why these areas would be distressed in the first place.

<sup>4</sup> I neglect here a widespread but politically naive position of many economists that policy should pursue efficiency objectives exclusively and take care of the problems of equity through transfer payments.

which are functionally rather than geographically defined. On the other hand, the legislative branch consists of representatives of territorially defined constituencies. Every piece of legislation, therefore, is scrutinized by the representative of each area to determine its local impact for his constituency. Thus, programs which do not appear on the surface to be regional are opposed, supported, or modified by men who are evaluating the regional consequences. This applies to tariff regulations, depreciation allowances, housing programs, welfare programs, agricultural subsidies, military expenditures, and all else. The processes involved are many, subtle, and varied. They include the production and dissemination of information as well as the uses of power and influence. Among the most visible manifestations are census studies, the support of research, public relations and propaganda, committee hearings, lobbying, and the trading of favors among legislators called pork-barrelling.

A sense of the actual scale of the regionalization of the United States from this point of view may be gained by considering that each member of the House of Representatives at present represents approximately 460,000 people. Until a few years ago the process of congressional reapportionment was extremely slow, so that urban population was under-represented while slower-growing or declining areas were over-represented. In consequence, the interests of the more backward areas tended to have disproportionate power<sup>5</sup>. In recent years a number of Supreme Court decisions have strengthened the « one man, one vote » principle and accelerated the pace of redistricting to conform with the changing patterns of population distribution. Although the immediate effect of this is to increase the representation of the growing and prosperous areas, a secondary effect has been to increase the urgency with which representatives from declining areas try to favor their economic growth, since the penalty of failure may be the elimination of their seat through redistricting. Thus, there has been no visible change in the « jobs to people » attitude of federal programs, and very few efforts have been made to facilitate the movement of people to where the jobs are.

The rhetoric commonly used in support of the « jobs to people » approach in the United States as in other countries, often makes use of a verbal trick which may be termed the geographic fallacy. The fallacy is to use the name of a territory, which is of course not sentient, for the population of that territory. Thus, outmigration of an economically overpopulated area may improve conditions both for those who leave and those who stay. But some may point to the diminishing population and diminishing regional product and say that the region is declining. This view ignores the mobility of people across regional boundaries. The Supreme Court, in its historic recent *Baker vs. Carr* decision, said: « Legislators represent people, not acres or trees. Legislators are elected

<sup>5</sup> This power is augmented by the tendency of these areas to elect the same man repeatedly. This gives considerable seniority to these Representatives, which places them in the most powerful committees.

by votes, not farms or cities or economic interests . . . The weight of a citizen's vote cannot be made to depend on where he lives ». The Court mirrors the eighteenth century rationalism of the Constitution. Similarly, most economists, I believe, would lay greater stress on policies to facilitate migration to job opportunities than does the U. S. Congress, especially as a long run solution. Low income and high unemployment may be viewed as the result of the imperfect mobility of labor as a production factor. If labor would move to where the returns to it are higher, wages and unemployment would be equalized throughout the economy<sup>6</sup>. In this view, economists are also the descendants of eighteenth century rationalism. Theirs is a world of individuals in a contract economy, much as the concept of government in the U.S. Constitution is based on the idea of social contract.

The geographic fallacy has little to recommend it, but it may be to some extent a poor statement of a point of view which economists neglect. What may be meant is that the rationalist solution may be at the expense, not of an *area*, but of a *people* or a *folk*. The social organization of distressed areas will be typically characterized to a lesser degree by contractual relations than the social organization of the more prosperous and modern areas. Rather, a greater role will be played by tradition, custom, complex networks of generalized obligations, and a more complete identification of the individual with his community. Thus, even if both those who leave and those who stay are materially better off as the result of the outmigration, this rationalistic solution may ignore psychological and cultural losses. It is, of course, easy to romanticize this type of consideration out of proportion, and in my opinion a great many areas are poor precisely because of their *Gemeinschaft*. Nonetheless, this quite real consideration, which would bolster the validity of the American approach to its distressed areas, is seldom articulated clearly and is always excluded from formal analysis.

Although, as discussed, the most significant federal programs are aimed at areas that meet qualifying criteria rather than at specific areas, there is a vast number of area-specific programs. Most of these are quite narrow, dealing with specific dams, harbors, and the like. However, the most ambitious of the area-specific programs are the newly created Regional Commissions. The first, the most active, and the best known of these is the Appalachia Regional Commission, established in 1965. The area defined by the Act has a 1960 population of 17,000,000, and covers all or portions of 13 states. The origins of the approach go back several decades<sup>7</sup>, and some of its features are reminiscent of an earlier departure, the

<sup>6</sup> This assumes that capital and entrepreneurship are more mobile and better informed than labor. In passing we may note a curious twist of current programs. Substantial outmigration, which serves to qualify an area as distressed, is regarded as evidence of local need rather than as successful adaption of the population to greater opportunities in other areas.

<sup>7</sup> See S.A. Levitan, *Federal Aid to Depressed Areas*, Baltimore: The Johns Hopkins Press, 1964.

Tennessee Valley Authority. However, the seminal moment was probably John F. Kennedy's visit to West Virginia during the 1960 presidential campaign. He was shocked by what he saw, and so were millions of Americans who, having no idea such conditions existed in America, saw it too through television. The establishment of the Commission took some years, and in the meantime the drain upon national resources of the Vietnamese conflict has limited expansion of the program, while the overwhelming urgency of the problem of interracial equity has cast the problem of interregional equity into a shadow. Nonetheless, substantial efforts continue in this direction, and occasional reports of conditions, such as a recent one documenting widespread malnutrition, revive a sense of importance. In the meantime, some features of theory and practice are of interest. No major studies of the work of the Commission are yet available, and I can only offer some remarks as an interested observer.

When President Johnson announced the formation of the Commission, he called it an important step toward a new « creative federalism ». The Commission consists of a Federal Co-chairman who is a personal representative of the President, and the Governors of the component states or their representatives, from among whom the other Co-chairman is named. The basic idea is the partnership of the Federal Government with the states for regional development, rather than a hierarchical relation. Thus, for instance, the major allocations will not be determined from Washington, but will be negotiated among the participating states which must themselves provide some of the development funds. Each state has produced a plan, and the collection of these constitutes the closest thing to an overall regional plan. Whether this structure can be effective over the long run or whether it will disperse efforts in compromises and half-measures remains to be seen. It does reflect, however, a widespread shift in American opinion toward greater local self-determination and a more direct participation by the smaller units in large decisions. Over the past several decades American conservatives had stressed local independence and the rights of states. In fact, some of the most conservative Republican candidates in recent years have run on a platform which amounts to a dismantling of much of the Federal Government. But American liberals had been centralists since at least the early days of the New Deal, advocating that one function after another be assumed by the Federal Government. Now the liberals increasingly call for more local autonomy for a variety of reasons ranging from a belief that greater and more direct participation is needed to restore to the individual a sense of control of his own destiny in a society which has grown too large and impersonal, to more technical arguments concerning local differences in subjective rates of time discount, length of feedback loops, and other issues which some economists are calling fiscal federalism. In the extreme, of course, there is the New Left which, while encompassing many divergent viewpoints (often within the same individual), often takes positions reminiscent of an earlier anarchistic syndicalism, in which the larger establishment is disassembled and voluntary associations guide their own

destinies and cooperate with their neighbors. This breadth of viewpoints in support for greater local self-determination is a new phenomenon in contemporary America. Its manifestations range from the demand by many ghetto residents to take over the police function within their communities, to widely supported suggestions that the Federal Government downgrade categorical programs and transfer lump sum payments to be spent at their discretion. It is uncertain where the balances will be struck between centralization and localism, but it is clear that there is a search for a new equilibrium, and that this will effect profoundly the philosophy and practice of regional development efforts.

The first years of work of the Commission provide an instance of a widespread and well publicized phenomenon in American government. This is the attempt to rationalize the process of making decisions and to quantify the measurement of as many variables as possible. The air is replete with the sound of cost-benefit analysis, cost-efficiency analysis, operations research, social indicators, and PPBS (Planning-Programming-Budgeting Systems). Although rationality must be welcome, there must be self-examination in the acceptance of these methods. In the first place, the increase in rationality is often an increase in technical rationality not easily translated into a layman's language. This tends to increase the power of the technocrat, and runs counter to the new winds of participatory democracy. It may be countered that the technician can simply present his findings neutrally to the people or their representatives for their decision, but it is by now well known that the cost of analysis, the difficulty of understanding the adequacy and limitations of conclusions, and the pace of events in real time weaken the effectiveness of the more diffuse popular body in its relations with the technical establishment. This conflict is avoided only if the technical contribution so increases control over destiny that both technicians and popular representatives increase their power.

But it is too easy to exaggerate the strength of the new techniques. An evaluative planning rationality depends on two things: on the ability to predict consequences and on the wisdom with which consequences can be judged. Neither is particularly advanced in the regional case although an important beginning is occurring. It is very difficult to predict for very open systems such as regions. What will be the consequences of improving the road access to a town? Will it attract jobs and improve the competitive position of the merchants in the town or will it make it more accessible to outside competition? If better education and worker training is provided, will it stimulate local economic activity or depress it, since it is the skilled and the educated who leave? Evaluation, too, is most difficult. Many questions remain to be solved in theory and many more are quite intractable in practice. Among the most difficult problems are dealing with the incidence of costs and benefits, taking account of non-monetary costs and benefits, discounting appropriately different time preferences, and determining shadow prices and opportunity costs properly.

The general difficulties of these methods increase in the case of regional analysis. For instance, operationally cost-benefit analysis makes



much use of the assumption that factor prices represent marginal factor costs, and that consumer demand prices reveal marginal utility conditions. This follows from classical theory in a market economy which is well integrated. But the difficulty with a great many depressed areas is that they are not economically or socially well integrated or particularly rational so that actual prices bear a less clear relation to marginality conditions. Similarly, the operational theory usually has to assume that the changes contemplated are not so large as to change these marginal values. Yet in the regional case, the change will often be large in relation to the base of departure, and this assumption is much weakened. Not all of the theoretical difficulties are so abstruse. For instance, proposals for Appalachia have stressed increases in tourism. If this increase is achieved, it would increase economic activity in the region. Most analysis counts this as a benefit, and so it would be from the point of view of Appalachia. But if the tourist from Chicago reduces his expenditures in Chicago by the amount he spends in Appalachia, the net result is no net increase in national product<sup>8</sup>. We have then that what is achieved is a geographic transfer of expenditure which would serve both the national objective of equity and the regional objective of growth, but not the national objective of growth. This simple distinction is ignored, and common practice counts tourist expenditures, increased by a regional multiplier, as simple benefits, i.e., additions to national product.

An examination of actual government investments in Appalachia is instructive. By far the greatest part of expenditures have gone into road construction. This is understandable since a great deal of that investment had already been programmed before the formation of the Commission and since powerful interests back the national road program<sup>9</sup>. On the other hand, it is an indication of the difficulty of transcending infrastructure projects, which merely set the stage for the economic activity which it is hoped they will bring about. The techniques which have just been discussed may be good for evaluating projects and their consequences, but they are very weak for generating them. Operationally, it is fairly easy to find an appropriate location for a particular industry, but it is extremely difficult to identify specific industries that can be attracted to or generated in a given area. Consequently, very little investment goes into directly productive capital, and a great deal into infrastructure.

Similarly, there is a commitment in principle to a policy of growth centers, called Growth Areas by the Appalachia Regional Commission and

<sup>8</sup> The analysis can be deepened by consideration of changes in the marginal propensity to consume, valuation of tourism not in terms of costs but of consumer surplus, etc., but these would only modify the argument.

<sup>9</sup> Other principal categories are construction of health facilities, land improvement and erosion control, vocational educational facilities, and supplements to various federal grant-in-aid programs.

<sup>10</sup> It must be recognized that population is not a particularly good measure, merely a convenient one. Urban scale is multidimensional. Similarly, the relevant minimum size will depend on local circumstances.

Economic Development Centers by the Economic Development Administration, which is in general charge of distressed area programs. The logic of these, at heart, is that economies of scale and externalities suggest that development had better be concentrated rather than dispersed throughout the area. However, beyond this appealing argument, the concept in the United States as elsewhere is weak both in theory and useful fact. How big should the centers be? Under some legislation, no center may be over one quarter million. Some of the designated centers have less than 5,000 population. Academic opinion, based on admittedly slender evidence, currently puts take-off size for urban areas in the United States in the range of 250,000 to 450,000<sup>10</sup>. But there is very little knowledge of the mechanisms or the parameters of scale effects and externalities. Thus, some versions suggest that the centers should be spaced at twice some maximum commuting radius, so that people will stay in their present homes and commute to these places for work. Other versions appear to conceive of substantial short-distance migrations to these centers, and imply that this localized population displacement is preferable to a long-distance migration out of the region. In actual fact, the American growth center strategy seems to be running into the well known problem of such a policy in democratic countries. Concentration of effort and investment in some localities means that others are slighted, and local political power makes it very difficult to hold to such restraint. There is therefore a tendency to a proliferation of centers which weakens their logic. Appalachia has some two hundred designated Growth Areas<sup>11</sup>, in the order of one per hundred thousand population.

Although there is increasing discussion of the need for a national urbanization policy, there is as yet no agreement what subjects it should deal with<sup>12</sup>. This discussion, however, manifests a new concern for the issue of the national system of cities, going beyond the traditional concern for the internal organization of urban areas. Thus, a policy of growth centers is a recognition that many economically distressed areas are organized into agglomerations which are too small. At the other extreme, Gotman's term megalopolis has entered common usage, and the national government is sponsoring economic and engineering studies of a high speed transportation link along the spine of the Eastern Seaboard megalopolis.

The classic pair of issues of national urbanization policies, new towns and stopping the growth of big cities, have received less attention in the United States than in many other countries. The current Housing Act has

<sup>11</sup> Each state designates its growth areas as part of its plan. Under the Economic Development Administration, each Economic Development District, which consists of several counties, designates one or more Economic Development Centers. Usually the largest settlement is so designated.

<sup>12</sup> Some idea of the status of the discussion may be gained from *National Growth and Its Distribution: Symposium on Communities of Tomorrow*, U. S. Government Printing Office, 1968. These proceedings are particularly interesting because they included not only academics and administrators but also the Vice President and several members of the Cabinet.

some mild provisions to facilitate the financing of new towns, which are privately built, and some, such as Reston and Columbia, have received considerable publicity. There is a great deal of talk in certain circles about big cities being too big and the need for a new town module with which to reurbanize America. The proponents of such policies include not only the usual architectonic planners and romantic humanistic utopians, but also some of the highest political figures, for whom the nostalgic vision of small-town America still has validity, scientists who visualize self-contained technological marvels, conservationists concerned about pollution, and large corporations which persuade themselves of the business opportunities of building new towns. There are only small beginnings of rational studies of the issues involved. At the moment, as in Europe and the developing countries, proof is by rhetoric and repetition, and no one can demonstrate that the big cities are too big (or that they are not) or that new towns make sense (or that they do not).

Although the internal structure of metropolitan areas is not our concern here, the problem of race has recently linked this issue with issues of regional planning and urbanization policies. In American metropolitan areas, the overwhelming majority of Negroes live in the central municipalities, and a substantial majority of whites live in the suburbs. Thus, some opposition to new towns is based on the likelihood of their resulting in further segregation by class and race, used as a means of escape by whites from urban areas which are increasingly disturbed by Negro numbers and militancy. At the same time, it has dawned on American consciousness that the Negro remains locked in the central ghetto while jobs continue to suburbanize, and that a strong factor in his unemployment and economic deprivation is his inability to reach jobs which may be ten to fifty kilometers away. With Euclidean logic, this realization has given rise to three types of proposals: (a) subsidies to bring jobs back to the central areas of the metropolis, (b) subsidies and legislation to enable the Negro to move to the suburbs, and (c) subsidies and other forms of encouragement to transportation links that will enable the Negro to remain in the center and reach jobs in the suburbs. Current efforts are still small, by comparison to the magnitude of the problem, but they pursue all three directions, and it is certain that more will be done in this issue.

Very recently an old argument has been advanced with new force and is receiving considerable attention<sup>13</sup>. This points to the condition of the Negro as a recent migrant from rural areas who is undergoing a particularly difficult process of acculturation and gaining acceptance only slowly. It is being suggested that improvement of urban conditions for the Negro is a two-edged proposition, and that efforts to provide more and better jobs for Negroes in the cities will increase Negro migration from rural areas perhaps to the extent of increasing the number of poor and

<sup>13</sup> John F. Kain is a principal writer on this subject. See, for instance, J. F. Kain and J. F. Persky, «The Ghetto, The Metropolis, and The Nation», Harvard Program on Regional and Urban Economics, Discussion Paper No. 30, 1968.

unemployed. Therefore, it is argued, although substantial efforts must be made in urban areas, a vigorous program should be undertaken for the development of the South, so as to provide opportunities there and thus diminish the propensity to migrate to the large urban areas of the rest of the country. This will give these areas more time and a better chance to absorb their Negro population.

This argument has an attractive symmetry and may be true in the long run. However, several counter arguments may be advanced. Given the present number of Negroes presently in the South and presently in the North, and the absolute size of the economies of the South and the rest of the country, the rate of growth that this policy would require of the Southern economy may be unrealistic, even with substantial federal help. Further, a policy of Southern development to slow the Negro migration to the rest of the nation will take time to discuss, adopt, and implement, and it is likely to be slow to take effect. The problem is vitally urgent, and the Negro leadership justly asks for equality NOW. Thus, such a long run policy does not take into account the reality of actual time and the reality of the American political situation, and must be considered as a background theme in the orchestration of efforts to meet the problem. Lastly, it must be noted that the South is a racist society almost entirely represented at the local, state, and federal levels by an archaic, almost feudal, power structure which is opposed to change if this brings with it changes in the structure of classes and races. It is extremely doubtful that this Southern establishment would be willing to accept policies that would change the fundamental structure of power, privilege, and racial relations. Similarly, there is anecdotal evidence that some of the Negro migration seeks to escape to the lesser racism of the North. It is doubtful that the Negro movement and their liberal allies would accept a policy to retain Negroes in the most racist part of the country.

Nonetheless, this type of argument represents an advance because it ties together into one system the development of large-scale regions, the flows of migrants, and the internal structure of metropolitan areas, so that changes in one factor bring about changes in the others. It points out that the distinctions of center and periphery at the national level are inverted within metropolitan areas, where it is the center that is poor and backward and the periphery that is prosperous and growing. At the same time, it brings into question, without easily discernible answers, whether the problems of the large urban areas in America are fundamentally urban problems or fundamentally racial problems. My opinion is that they are the latter, and that it is necessary to incorporate into our conceptual system not only the interrelations between the intra-urban and the inter-regional scales, but also the interrelations between social and economic sectors as holders of economic, political, and moral power. Let me point out, for instance, that much of the Negro leadership currently opposes integration and advocates economic development within the ghetto, including ownership of the means of production, as a way of improving the social and economic status of black people by increasing their solidarity

and their ability to move upward as a class. Liberal economics and sociology are having great difficulty in accepting this more organic view of social structure.

#### THEORY AND TECHNIQUE IN RELATION TO REGIONAL PLANNING

Abstraction and generalization of theory and technique is a difficult and dangerous matter, for it avoids the concreteness and detail that makes these valuable. True generalization is rare and usually represents a significant contribution. I have no such pretensions here. The discussion limits itself to some general observations concerning the state of the art in the United States, and the relation of techniques to policies and programs.

Applied regional economics, as distinct from location theory, consists primarily of models based on the concept of multipliers. These models have advanced significantly in the past decade, and today they range from the simple economic base model to elaborate social accounting and econometric simulations. Among the improvements and elaborations are the disaggregation of populations into relevant sub-categories, occasional consideration of price effects, more sensitive distinctions between stocks and flows, and, especially in simulation models, the incorporation of variables whose metric is not cognate with the dollars, tons, or workers of the basic models. Particularly interesting are current attempts to meld together demographic and economic models. Demographic models have projected population trends, but left the economic question unanswered. Economic models have projected jobs, and assumed that population would adjust itself to them. Theory and recent empirical work show that there is significant tendency both for jobs to follow people and for people to follow jobs, and that both types of cross-elasticity are significant. Such advances indicate that regional models are still improving, and that they will probably become more useful instruments. Still, the basic question in many cases is not what will be the long range growth of a regional economy or what will be the multiplier of a particular increase in the export sector. In the case of distressed regions, the pressing question tends to be how on earth to identify and attract economic activity to the area, and regional economics as yet can contribute very little to this question<sup>14</sup>.

A curious mismatch may be noted between the most commonly used regional models and some of the principal ideas of development strategy. For instance, the vast majority of regional models conceive of a region as a bounded space of no internal spatial dimensions. That is to say, the region is a point-economy, and a person or a dollar is clearly inside or outside of it at any moment in time. Further, the models generally assume constant factor proportions, constant returns to scale, and the like. Yet the strategies of regional development base a great deal of their thinking

<sup>14</sup> By contrast, current Soviet practice makes extensive use of regional models of the input-output family to determine industrial location.

on the concepts of continuous space and of effects of scale. Consider the strategy of growth poles. It assumes that a pole affects the territory around it, and that the strength of that effect declines with distance. A particular locality may be affected by more than one pole. This is a view of space as continuous and without hard regional boundaries. At the same time, most formulations of growth poles stress the concept of critical size, below which there is not the scale to support certain lumpy activities, to permit functional specialization and efficiency, and to achieve the reverberation of externalities that will set off self-sustaining growth. Contrast this, for instance, with the assumptions implicit in the use of technical coefficients in an input-output model: here all that matters are the relative numbers and an economy may double its size with no change in its industrial structure. While central place theory is closer to the conceptualization of space of a growth pole strategy, it has been elaborated on the basis of the marketing of consumption goods and the assembly of agricultural produce and at present lacks an operational theory or a convenient mathematics to study questions relating to the location of other economic activities and their interaction through externalities within a center.

There is also a new school of geographers, who apply advanced statistics such as factor analysis and canonical correlation to the problems of regionalization and regional flows. Geographers are beginning to shed their traditional mistrust of abstraction and highly interesting theoretical contributions are being made, although perhaps these do not match the new statistical power. Among the developments are attempts to identify the basic dimensionality of socio-economic structures in space, applications of graph and set theory and spectral analysis, and uses of communications and general systems theory.

Applied and scientific work has been aided by several factors. Censal information is available in greater detail by fine geographic divisions, often directly in computer tapes. Computers are more powerful, easier to use, and developing new capacities such as computer graphics. National interest in these topics has channelled considerable funds to applied work, and there begins to be more investment in support of scholarly work and students. Some years ago, the scholars who specialized in localized concerns were at the bottom of the intellectual hierarchy and lowest in prestige in such fields as economics, political science, and sociology. The combination of funds, the intellectual stimulation of scientific advances, and the vital importance of regional problems (especially urban) has begun to attract the main component for continued development of the subject: intelligent men<sup>15</sup>. Today there is a free-masonry of planners, economists, geographers, sociologists, systems analysts, statisticians, and others with common attitudes, common technical language, overlapping readership of journals and attendance at conferences. Most of the significant work, how-

<sup>15</sup> Aerospace and other scientifically-based industries are exploring the transferability of competence in systems analysis to urban and regional issues although thus far they have not met with distinguished success in this experiment.

ever, is not being done by the traditional interdisciplinary formula (« take an economist, a sociologist, an architect, blend and pour ») but by men who do not feel constrained by the traditional disciplinary boundaries. The disciplinary labels become arbitrary when applied to many of these men and their work.

The field of evaluative techniques is less developed than those of analysis and projection, although, of course, the distinction is not a sharp one. We have mentioned the great interest in cost-benefit analysis, social indicators, budget programming, efficiency analysis and the like, and some of the practical and theoretical difficulties involved in their application. But even after taking account of faddishness and inflated formalism, this type of work is so clearly needed and its basic premises of rationalizing and quantifying need and performance are so undeniable that one may predict with confidence that its use will grow in frequency, importance, and, hopefully, quality. It must be remembered that Robert McNamara did not arrive in Washington as Secretary of Defense until 1961, and that it was not until 1965 that President Lyndon Johnson directed the agencies of the Federal Government to adopt the techniques of Planning-Programming-Budgeting Systems. The United States is committing money and brains to this subject, and, as theory continues to sort itself out and behavioral parameters are identified and measured greater precision, we may expect a useful routinizing of applications so that they may be performed by ordinary technicians as a matter of course rather than having each occasion be a difficult, uncertain, and expensive attempt to invent something new.

Although the United States is advancing rapidly in what has come to be called hard techniques, other aspects of regional studies have lagged by comparison. The non-quantifiable aspects have not been developed at the same rate or in a systematic, cumulative fashion. There is indeed a wealth of literature, much of it perceptive, some of it organized, on social, political, and institutional aspects, but this river of discourse flows in a separate bed from that of more formal quantified science. It appears that this has brought about a curious circumstance. The hard techniques have become an important factors in determining private and public actions, and regional scientists begin to find themselves, as their work becomes important and expensive, both as actors who influence the phenomena they study and as participants who are themselves affected. For instance, it has been remarked repeatedly that the reportage of such complex applied regional models as have been done is very sketchy, and that this makes them hard to understand, hard to judge, and wasteful in that little is learned from prior failures and successes. It seems to me that this lack of reportage comes not so much from the laziness or modesty of the scholars involved as from a lack of recognition of the interdependence of technical work with its institutional matrix. We have not developed styles for reporting these aspects, and therefore the reports are either confusing or misleading, presenting the work as if it were done in scientific isolation. There is great need to study the relations

of the technicians to their clients in the process of analyzing and planning, and, later, in how the plans are used in the political process of decision and the actual process of implementation<sup>16</sup>.

There is in the United States no commonly accepted theory of planning. Indeed, there is some residue, although mostly a verbal one, of feeling against the concept of planning, particularly at the federal level, stemming from an association of such plans with command societies and from an opposition to restrictions upon individual freedom on decision. Nonetheless, a great deal of explicit regional planning goes on, and we have argued that territorial representation in the legislative branch results in implicit regional planning. Since decisions for regional development take place in a complex and shifting polity, hierarchical relations are weak, and the plans are more akin to an action painting than to a carefully composed canvas. The spirit of this type of planning is caught by Lindblom's theory of disjointed incrementalism<sup>17</sup> which argues that social mechanisms are too complex for wholistic plans to have much validity, and that decisions should be taken incrementally, by small steps. Similarly, the French concepts of indicative planning strike a responsive chord with many Americans, since it reflects the American approach of planning by making information available, by inducements and negotiations. In fact, the theorizing of the early New Deal by men such as R. Tugwell and Col. H. Johnson bears a striking resemblance to contemporary French descriptions of the process of *Le Plan*. Under these conditions it is not surprising that there should be little emphasis on master plans, and that such plans as exist are regarded as rolling plans and not taken too seriously. Rather, there is emphasis on analysis and projection, on voluntary coordination, and upon techniques such as cost-benefit which are used to evaluate marginal increments (concrete actions) on the basis of generalized goals and criteria. There is, of course, the danger that hidden interdependences may result in unanticipated consequences, and that the whole may be either more or less than the sum of the parts. Whether the prevalent American style of planning without master plans is better or worse than the frequently advocated model of planning by comprehensive plans is at this time a matter of opinion and a subject of widespread debate.

<sup>16</sup> The most interesting discussion of this aspect is to be found in T. C. Schelling, « PPBS and Foreign Affairs », *The Public Interest*, Spring, 1968. Schelling is discussing the application of these techniques for national policy with respect to other nations and the regions of the world, but most of what he says is of equal relevance for regions which are internal to the nation as for external regions. See also, A. Wildawsky, « The Political Economy of Efficiency: Cost-Benefit Analysis, Systems Analysis, and Program Budgeting », *Public Administration Review*, Dec. 1966.

<sup>17</sup> See D. Baybrooke and C. Lindblom, *A Strategy for Decision*, New York, 1963, and later writings by Lindblom. See also W. Alonso, « Predicting Best With Imperfect Data », *Journal of the American Institute of Planners*, July, 1968, for some arguments about the mathematical unreliability of complex arguments.



---

**Part V Bibliographical Essay**

---

John Friedmann

Until the late 1950s, most people understood regional planning to mean the purposeful development of a region's natural resources. Such regions were generally defined as river basins, and the principal concern was therefore with the development of water (for the multiple uses of power, irrigation, flood control, navigation, and recreation) and land (for regulating runoff, controlling erosion, and stimulating the production of agriculture and livestock). Urban-industrial development was by-and-large regarded as only an auxiliary element in the buildup of a region's economy. The model of this comprehensive approach to the harnessing of a region's resources was, and still is, the Tennessee Valley Authority—and its successors throughout the world in Colombia (Cauca Valley), Ghana (Volta River), India (Damodar Valley), Egypt (Aswân), and elsewhere. As significant as this concept continues to be, however, it did not readily lend itself to systematic treatment: the regional approach suggested a collaborative effort by groups of different specialists, such as engineers, agronomists, economists, and administrators, but regional planners as such were not among them. It was only subsequent to 1955 that a general concept of regional planning came gradually to be accepted as the coordinative framework for a variety of development activities at subnational levels. The revival of interest in location economics and spatial systems analysis that followed publication of Walter Isard's initial pathbreaking contributions to the literature suggested that the *ordering of economic activities and human settlements in physical space* might become the central focus of a new discipline of regional planning. *Regional Development and Planning* (1964), edited by John Friedmann and William Alonso, was an attempt to lay the bases of this discipline. Its theoretical thrust was made possible by confining attention to a fairly narrow range of concerns—location and spatial organization—but these proved to be extraordinarily powerful concepts that touched, directly or indirectly, on a wide array of vital policy concerns. Some of these will be discussed in this chapter.

A decade's work has yielded a rich harvest. It has not only deepened the theoretical foundations of regional planning but has extended knowledge drawn from policy analysis and the evaluation of actual experiences with regional planning in a variety of national settings. A brief review of progress in theory, policy analysis, and experience is attempted in the following three sections.

In the fourth section, attention turns to the principal tasks of regional planning. A convergence upon two distinct problem sets will be noticed. In

industrially advanced and postindustrial societies, regional planning is primarily concerned with problems of depressed areas, regional adjustments to locational shifts in major economic activities, and the structural and environmental problems of expanding urban regions. In developing countries, the focus is more on the regional coordination of investment programs, decentralization of decision-making, alleviation of regional imbalances, and the spatial integration of the national society. It is obvious from even this brief summary that regional planning addresses itself to significant issues in societal development. It is all the more curious, therefore, to observe the halfhearted efforts at regional planning in the United States. Some reasons for this anomalous behavior will be advanced.

The final section casts a prospective look into the future. Several promising lines of inquiry are suggested as an agenda for research for the coming decade. If the discipline of regional planning is to remain fresh and alive, it is clear that this volume will have to be rendered obsolete by 1984, just as our best thinking today cancels, or at least improves upon, much of what we thought to be true a mere ten years ago.

#### Progress in Theoretical Knowledge.

A textbook codifies existing knowledge in a field and presents it in a readable and systematic way. At least six textbooks and as many collections of readings on regional economics, policy, and planning have been published since 1964.<sup>1</sup> It is typical for this increasingly international field that of the six textbook authors, only two are American; the others are German, English, Norwegian, and Dutch. (The Norwegian contribution was written primarily for Indian students and practitioners.) The essay collections are similarly directed at various national markets: England, Holland, Italy, Venezuela, and the United States. National diversity, however, resolves into only two major theoretical orientations that distinguish among these texts. One group of authors (Hugh O. Nourse, Horst Siebert, and to some extent Harry W.

<sup>1</sup>Texts: Hugh O. Nourse, *Regional Economics: A Study in the Economic Structure, Stability, and Growth of Regions* (New York: McGraw-Hill, 1968); Harry W. Richardson, *Regional Economics: Location Theory, Urban Structure, and Regional Change* (New York: Praeger, 1969); Horst Siebert, *Regional Economic Growth: Theory and Policy* (Scranton: International Textbook Co., 1969); Tormod Hermansen, *Spatial Organization and Economic Development: The Scope and Task of Spatial Planning* (Mysore: University of Mysore, Institute of Development Studies, 1971); Jos G. M. Hilhorst, *Regional Planning: A Systems Approach* (Rotterdam: Rotterdam University Press, 1971); Edgar M. Hoover, *An Introduction to Regional Economics* (New York: Alfred A. Knopf, 1971).

Collections of readings: L. Needleman, ed., *Regional Analysis: Selected Readings* (London: Penguin Books, 1968); Robert D. Dean, William H. Leahy, and David L. McKee, eds., *Spatial Economic Theory* (New York: The Free Press, 1970); David L. McKee, Robert D. Dean, and William H. Leahy, eds., *Regional Economics: Theory and Practice* (New York: The Free Press, 1970); Alfredo Testi, ed., *Sviluppo e pianificazione regionale: le teorie e le politiche* (Turin: G. Einaudi, 1970); Leo H. Klaassen, ed., *Regionale economie: het ruimtelijke element in de economie* (Groningen: Wolters-Nordhoff N.V., 1972); Julio César Funes, ed., *La ciudad y la región para el desarrollo* (Caracas: Comisión de administración pública de Venezuela, 1972).

Richardson) adopt an explicit *nonspatial* approach to their subject. Their models employ a concept of discrete regions, with the result that the national economy appears as the aggregate of two or more linked subnational or regional economies each with its own characteristics. The other authors (Jos G. M. Hilhorst, Tormod Hermansen, Edgar M. Hoover, and again in part Richardson) may be called *spatial* theorists; they operate with a concept of a continuous development surface that exhibits certain characteristics of spatial structure. As a result, these authors are likely to emphasize such phenomena as urban size hierarchies (nodes in a system of economic relationships extended over the landscape), the polarization in space of economic production and settlement, core-periphery relationships, and urbanization processes. Any point on a development surface is therefore situated in a structured field of forces that, to a large extent, determines the economic growth potential at that location. The spatial approach has emerged as the dominant theoretical orientation in regional planning and, of necessity, leans strongly on analytical geography as its root discipline. The nonspatial approach, on the other hand, draws primarily on the analytical concepts of economics. During the past decade, it has been the spatial approach, however, that has scored the more substantial advances in theory.<sup>2</sup> Of particular interest to regional planners are four major research frontiers: spatial diffusion of innovations, geography of development, structure and growth of urban systems, and polarized development.

**Spatial diffusion of innovations.** The origin of spatial diffusion research is deeply indebted to the Swedish geographer, Torstén Hägerstrand of Lund University.<sup>3</sup> Early diffusion studies were almost wholly concerned with measuring the rate and direction of the diffusion of discrete innovations, such as a product, a technological improvement, or a particular organizational form, and with describing this process as it unfolded in geographic space, together with the intervening variables and the underlying processes of communication.<sup>4</sup> But soon the work branched out in a number of directions that proved to be of exceptional importance to regional planning. Gunnar Törnqvist<sup>5</sup>

<sup>2</sup>A series of textbooks summarize the current state of knowledge in spatial organization and development, including Peter Haggett, *Locational Analysis in Human Geography* (London: Edward Arnold, 1965); Brian J. L. Berry, *Geography of Market Centers and Retail Distribution* (Englewood Cliffs: Prentice-Hall, 1970); Richard L. Morrill, *The Spatial Organization of Society* (Belmont, Calif.: Wadsworth Publishing Co., 1970); Ronald Abler, John S. Adams, and Peter Gould, *Spatial Organization: The Geographer's View of the World* (Englewood Cliffs: Prentice-Hall, 1972); Peter E. Lloyd and Peter Dicken, *Location in Space: A Theoretical Approach to Economic Geography* (New York: Harper & Row, 1972).

<sup>3</sup>Torstén Hägerstrand, *Innovation Diffusion as a Spatial Process*, trans. Allan Pred (Chicago: University of Chicago Press, 1967).

<sup>4</sup>A concise summary of this early phase appears in Peter Gould, *Spatial Diffusion*, Commission on College Geography, Resource Paper, no. 4 (Washington: Association of American Geographers, 1969); see also Lawrence A. Brown, *Diffusion Processes and Location: A Conceptual Framework and Bibliography*, Bibliography Series, no. 4 (Philadelphia: Regional Science Research Institute, 1968).

<sup>5</sup>Gunnar Törnqvist, *Contact Systems and Regional Development*, Lund Studies in Geography, Ser. B., Human Geography, no. 35 (Lund: C. W. K. Gleerup, 1970).

and B. Thorngren<sup>6</sup> undertook detailed studies of contact networks in production, business, and government, developing generalizations concerning the reach of these networks and their implications for location decisions. Allan Pred<sup>7</sup> showed how these contact networks underlie the process of spatial diffusion, while Poul Ove Pedersen<sup>8</sup> was able to relate the diffusion of successive waves of innovations through an urban system to the differential growth of cities and, by implication, of urban-dominated regional economies. **Geography of development.** Another group of geographers undertook to study the results of earlier diffusion processes by mapping the spatial distribution of selected indices of social and economic development. Brian Berry<sup>9</sup> devised a relatively simple method of measuring these indices along a series of traverses varying in length and passing through two or more cities in the United States. The graphs he obtained from these measurements clearly revealed the influence of cities in the landscape of economic development; in general, positive measures of development came to a peak in cities (and tended to be higher for large than for small cities) and declined with distance from these centers to the troughs of the "inter-metropolitan periphery."<sup>10</sup> A different methodology was applied to the study of "modernization" in Africa. The pioneer in this line of research was Edward Soja, whose empirical work centered in Kenya.<sup>11</sup> Subsequently, Peter Gould undertook a study of modernization surfaces for Tanzania,<sup>12</sup> while J. Barry Riddell did research on the relation between innovation diffusion and the evolution of modernization surfaces in Sierra Leone.<sup>13</sup> All of these studies revealed the existence of strong regularities in the data both for the topography of the surfaces with their

<sup>6</sup>B. Thorngren, "How Do Contact Systems Affect Regional Development?" *Environment and Planning*, vol. 2 (1970): 409-428.

<sup>7</sup>Allan Pred, "Urban Systems Development and the Long-Distance Flow of Information through Preelectronic U.S. Newspapers," *Economic Geography*, vol. 47, no. 4 (October 1971): 498-525; idem, *Information Circulation and the Process of Urban Growth: The U.S. System of Cities, 1790-1840* (Cambridge: Harvard University Press, 1973).

<sup>8</sup>Poul Ove Pedersen, "Innovation Diffusion within and between National Urban Systems," *Geographical Analysis*, vol. 2 (1970): 203-254; idem, *Urban-Regional Development in South America: A Process of Diffusion and Integration* (Paris: Mouton, 1974).

<sup>9</sup>Brian J. L. Berry, *Spatial Organization and Levels of Welfare: Degree of Metropolitan Labor Market Participation as a Variable in Economic Development* (Washington: Economic Development Administration, 1967).

<sup>10</sup>For the concept of the "inter-metropolitan periphery," see John Friedmann and John Miller, "The Urban Field," *Journal of the American Institute of Planners*, vol. 31, no. 4 (November 1965): 312-320.

<sup>11</sup>Edward Soja, *The Geography of Modernization in Kenya* (Syracuse: Syracuse University Press, 1968); Edward Soja and Richard J. Tobin, "The Geography of Modernization: Paths, Patterns, and Processes of Spatial Change in Developing Countries," in Ronald Brunner and Gary Brewer, eds., *A Policy Approach to the Study of Political Development and Change* (New York: The Free Press, 1974).

<sup>12</sup>Peter R. Gould, "Tanzania 1920-63: The Spatial Impress of the Modernization Process," *World Politics*, vol. 22, no. 2 (January 1970): 149-170, reprinted as chapter 10 in this volume.

<sup>13</sup>J. Barry Riddell, *The Spatial Dynamics of Modernization in Sierra-Leone: Structure, Diffusion, and Response* (Evanston: Northwestern University Press, 1970).

pronounced tendency to polarization and "corridor" development<sup>14</sup> and for changes in the form of these surfaces with progressive development in the direction of greater spatial integration. Both urbanization and distance effects could be isolated. Particularly intriguing is the similarity between the observed progression of modernization surfaces for a country like Tanzania and Jeffrey Williamson's thesis that, over the course of a country's economic development, regional income differences tend to move through a regular cycle from relative equality to growing divergence and ultimate convergence of incomes.<sup>15</sup>

**Structure and growth of urban systems.** As early as ten years ago, we knew that cities arranged themselves into hierarchical systems, and that the growth of these systems conformed to certain statistical laws.<sup>16</sup> But, especially in a context of development, many crucial questions remained unanswered. A detailed discussion of the research addressed to these questions is not possible here. Instead, I shall simply identify the principal questions together with references to the most important research related to them.

What is the relation between a country's level of urbanization and the percentage of its labor force in manufacturing industry? (N. V. Sovani<sup>17</sup> and David Kamerschen.<sup>18</sup>)

What consequences does an accelerated rate of urban growth have for social, economic, and political changes within a given country? (John Friedmann.<sup>19</sup>)

Under what conditions does primacy arise in the size distribution of cities? And what is the meaning of a rank-size distribution of urban population size? (Arnold Linsky,<sup>20</sup> César Vapnarsky,<sup>21</sup> Berry,<sup>22</sup> Salah El-Shakhs.<sup>23</sup>)

<sup>14</sup>C. F. J. Whebell, "Corridors: A Theory of Urban Systems," *Annals of the Association of American Geographers*, vol. 59, no. 1 (March 1969): 1-26.

<sup>15</sup>Jeffrey G. Williamson, "Regional Inequality and the Process of National Development: A Description of the Patterns," *Economic Development and Cultural Change*, vol. 13, no. 4, pt. 2 (July 1965): 3-45, reprinted as chapter 7 in this volume.

<sup>16</sup>Rutledge Vining, "An Outline of a Stochastic Model for the Study of the Spatial Structure and Development of a Human Population System," *Papers of the Regional Science Association*, vol. 13 (1964): 15-41, reprinted as chapter 1 in this volume. See also Brian J. L. Berry, "Cities as Systems within Systems of Cities," in John Friedmann and William Alonso, eds., *Regional Development and Planning: A Reader* (Cambridge: Press, 1964), chapter 6.

<sup>17</sup>N. V. Sovani, "The Analysis of 'Over-Urbanization,'" *Economic Development and Cultural Change*, vol. 12, no. 2 (January 1964): 113-122, reprinted as chapter 18 in this volume.

<sup>18</sup>David R. Kamerschen, "Further Analysis of Over-Urbanization," *Economic Development and Cultural Change*, vol. 17, no. 2 (January 1969): 235-253.

<sup>19</sup>John Friedmann, *Urbanization, Planning, and National Development* (Beverly Hills: Sage Publications, 1973), chapter 5.

<sup>20</sup>Arnold S. Linsky, "Some Generalizations concerning Primate Cities," *Annals of the Association of American Geographers*, vol. 55 (1965): 506-513.

<sup>21</sup>César A. Vapnarsky, "On Rank-Size Distribution of Cities: An Ecological Approach," *Economic Development and Cultural Change*, vol. 17, no. 4 (July 1969): 584-595.

<sup>22</sup>Brian J. L. Berry, "City Size and Economic Development: Conceptual Synthesis and Policy Problems," in Leo Jacobson and Ved Prakash, eds., *South and Southeast Asia Urban Affairs Annual*, vol. 1, *Urbanization and National Development* (Beverly Hills: Sage Pub-



What accounts for the observed stability in the rank-size ordering of cities? (J. R. Lasuén.<sup>24</sup>)

What is the relation between the function and size of cities under conditions of expanding urban employment? (Thomas Stanback and Richard Knight.<sup>25</sup>)

What influence does the spatial distribution of power have on the development of urban systems? (Anibal Quijano,<sup>26</sup> Friedmann.<sup>27</sup>)

**Polarized development.** The theory of polarized development is essentially a theory about the role of power relations in spatial development. Originally formulated as a nonspatial economic theory that provided an elaborate rationale for the concentration of economic power in a few technologically advanced industries tied into the rest of the economy through an elaborate system of input-output linkages,<sup>28</sup> it was soon "translated" into spatial terms by French and Belgian planners, notably J. R. Boudeville,<sup>29</sup> who discovered in the theory of polarized development a normative basis for their own work in *aménagement du territoire*. Policy applications quickly outpaced theoretical explanations, however, and two major reviews of the early literature by Niles Hansen<sup>30</sup> and D. F. Darwent<sup>31</sup> pointed to its grossly inadequate conceptualization and empirical foundations. Interest in spatial polarization nevertheless continued to be high, as the facts of geographic concentration in production, population, and employment seemed to be indisputably confirmed. The problem was to explain this cumulative concentration and, beyond this, to study the evolving relationship of an industrial "pole" with the remaining sectors of the space economy. In a book on regional policy in Venezuela, Friedmann<sup>32</sup> proposed a core-periphery model

lications, 1971).

<sup>23</sup>Salah El-Shakhs, "Development, Primacy, and System of Cities," *The Journal of Development Areas*, vol. 7, no. 1 (October 1972): 11-36.

<sup>24</sup>J. R. Lasuén, "Multi-Regional Economic Development: An Open Systems Approach," in Torston Hägerstrand and Antoni R. Kuklinski, eds., *Information Systems for Regional Development: A Seminar*. General Papers, Lund Studies in Geography, Ser. B., Human Geography, no. 37 (Lund: C. W. K. Gleerup, 1971).

<sup>25</sup>Thomas M. Stanback, Jr., and Richard V. Knight, *The Metropolitan Economy: The Process of Employment Expansion* (New York: Columbia University Press, 1970).

<sup>26</sup>Anibal Quijano, "Dependencia, cambio social y urbanización en Latino-américa," *Cuadernos de desarrollo urbano-regional* (Universidad católica de Chile), no. 6 (March 1968): 3-48; see also "The Urbanization of Society in Latin America," *Economic Bulletin for Latin America* (Economic Commission for Latin America), vol. 13, no. 2 (1968): 76-93.

<sup>27</sup>John Friedmann, "The Spatial Organization of Power in the Development of Urban Systems," *Development and Change*, vol. 4, no. 3 (1972-73): 12-50, reprinted as chapter 11 in this volume.

<sup>28</sup>José Luis Coraggio, "Hacia una revisión de la teoría de los polos de desarrollo," *Revista latinoamericana de estudios urbano regionales*, vol. 2, no. 4 (March 1972): 25-40.

<sup>29</sup>J. R. Boudeville, *Problems of Regional Economic Planning* (Edinburgh: At the University Press, 1966).

<sup>30</sup>Niles M. Hansen, "Development Pole Theory in a Regional Context," *Kyklos*, vol. 20 (1967): 709-725.

<sup>31</sup>D. F. Darwent, "Growth Poles and Growth Centers in Regional Planning: A Review," *Environment and Planning*, vol. 1, no. 1 (1969): 5-31, reprinted as chapter 25 in this volume.

<sup>32</sup>John Friedmann, *Regional Development Policy: A Case Study of Venezuela* (Cambridge:

that seemed to account for many of the relevant facts and challenged traditional assumptions concerning spatial equilibrium. This model was subsequently generalized in a theory of polarized development that attempted to integrate what was then known about innovation diffusion, urban system behavior, and interregional patterns of conflict.<sup>33</sup> A volume edited by Hansen in 1972 brought both this and several other contributions together, in particular the work of J. R. Lasuén, Morgan Thomas, and Tormod Hermansen.<sup>34</sup>

### Progress in Policy Analysis

A second major body of literature defining the field of regional planning covers normatively oriented or policy studies. These take their cue from the *problematique* of planning practice and attempt to put potential interventions in ongoing spatial processes on a "rational" basis. Five major policy issues have been and continue to be investigated by an international group of scholars: the question of optimum city size and size hierarchies, national urban growth policies, growth center policies, location policy, and subnational policies for full employment.

**Optimum city size and size hierarchies.** Is there a population size of cities that in some specified sense may be called "best" or "optimal"? And, further, is there an optimal arrangement of city sizes in a system of cities, such as a statistical log-normal distribution? Claims concerning the existence of such optima abound in the literature. A population size range of 250,000 to 500,000 is frequently mentioned as "optimal" based on either an assumption increasing marginal social costs of further additions to the population<sup>35</sup> or on a notion of threshold size beyond which further growth becomes self-sustaining.<sup>36</sup> Recent empirical work, however, has seriously challenged these claims (Alonso,<sup>37</sup> Richardson,<sup>38</sup> and Keichi Mera<sup>39</sup>). Data on urban productivity tended to show that, although the marginal social costs of urban

MIT Press, 1966).

<sup>33</sup>Friedmann, *Urbanization, Planning, and National Development*, chapter 3, "A General Theory of Polarized Development." This paper was widely circulated in mimeographed form from 1969 on.

<sup>34</sup>Niles M. Hansen, ed., *Growth Centers in Regional Economic Development* (New York: The Free Press, 1972). In particular, see J. R. Lasuén, "On Growth Poles"; Morgan D. Thomas, "Growth Pole Theory: An Examination of Some of Its Basic Concepts"; and Tormod Hermansen, "Development Poles and Related Theories: A Synoptic Review."

<sup>35</sup>G. M. Neutze, *Economic Policy and the Size of Cities* (Canberra: Australian National University, 1965); William Alonso, "The Question of City Size and National Policy," in Rolf Funck, ed., *Recent Developments in Regional Science* (London: Pion, 1972).

<sup>36</sup>Brian J. L. Berry, *Metropolitan Area Definition: A Re-evaluation of Concept and Statistical Practice*, Working Paper no. 28, revised (U.S. Department of Commerce, Bureau of the Census, July 1969), p. 25.

<sup>37</sup>Alonso, "Question of City Size and National Policy."

<sup>38</sup>Harry W. Richardson, "Optimality in City Size, Systems of Cities, and Urban Policy: A Sceptic's View," *Urban Studies*, vol. 9, no. 1 (February 1972): 29-48. For a more complete statement, see idem, *The Economics of Urban Size* (Lexington, Mass.: Saxon House/Lexington Books, 1973).

<sup>39</sup>Keichi Mera, "On the Urban Agglomeration and Economic Efficiency," *Economic Development and Cultural Change*, vol. 21, no. 2 (January 1973): 309-324.

growth might continue to rise beyond the assumed optimal size, the marginal productivity of labor rose even faster, quite possibly into the range of metropolitan regions with a population of several million. On the score of city size distributions, the evidence was less clear (see footnote 38), but at least it challenged the easy conclusion that a rank-size distribution may be regarded as the desirable pattern toward which urbanization policies should be unambiguously directed.

**National urban growth policies.** In recent years, questions have been raised concerning the spatial distribution of urban populations in a number of industrially advanced countries. This interest was occasioned by several phenomena that appeared, at least to the casual eye, to be related: the continuing drift of population from the interior, less urbanized regions of the United States toward "megalopolitan" coastal regions; growing preoccupation with environmental pollution; high densities, congestion, and a rising crime rate in large cities; and persistent poverty in large regions of primary production. The objective of most urban growth policies was to stem the further expansion of large cities and to decentralize population into new towns on the edge of the metropolis and into "growth centers" located in declining or stagnating "interior" regions.<sup>40</sup>

This set of policy issues gave occasion to a number of thoughtful articles that examined both the objectives and feasibility of national policies for urban decentralization. Among the more significant contributions are the rather pessimistic analyses of Anthony Downs,<sup>41</sup> Lowdon Wingo, Jr.,<sup>42</sup> and Wilbur Thompson.<sup>43</sup> One of the major contributors to this debate was William Alonso. In a series of closely reasoned essays, he demolished the usual rationales advanced in support of a massive national program for the construction of new towns,<sup>44</sup> provided evidence on the prevalence of what he called spontaneous growth centers in the United States,<sup>45</sup> and seriously questioned the feasibility of an explicit urban growth policy, given the unanticipated urbanization effects of most governmental policies directed at other objectives.<sup>46</sup> Chiefly as a result of mounting scepticism among policy

<sup>40</sup>National Goals Research Staff, *Towards Balanced Growth: Quantity with Quality* (Washington: Government Printing Office, 1970). See also John Friedmann, "The Feasibility of a National Settlement Policy," *Growth and Change*, vol. 2, no. 2 (April 1971): 18-21.

<sup>41</sup>Anthony Downs, "Alternative Forms of Future Urban Growth in the United States," *Journal of the American Institute of Planners*, vol. 36, no. 3 (January 1970): 3-11.

<sup>42</sup>Lowdon Wingo, Jr., "Issues in a National Urban Development Strategy for the United States," *Urban Studies*, vol. 9, no. 1 (February 1972): 3-27.

<sup>43</sup>Wilbur Thompson, "The National System of Cities as an Object of Public Policy," *Urban Studies*, vol. 9, no. 1 (February 1972): 99-117.

<sup>44</sup>William Alonso, "The Mirage of New Towns," *The Public Interest*, no. 19 (spring 1970): 3-17.

<sup>45</sup>William Alonso and Elliott Medrich, "Spontaneous Growth Centers in Twentieth-Century American Urbanization," in Hansen, *Growth Centers*, pp. 229-265.

<sup>46</sup>William Alonso, "Problems, Purposes, and Implicit Policies for a National Strategy of Urbanization," in U.S. Commission on Population Growth and the American Future, *Research Reports*, vol. 5, *Population Distribution and Policy*, ed. Sara Mills Mazie (Washington: Government Printing Office, 1972), chapter 3, reprinted as chapter 29 in this volume.

analysts, the government has quietly dropped a national urban growth policy from the national agenda.

**Growth center policies.** As mentioned earlier, the theory of polarized development all too quickly spilled over into normative applications. The amount of confusion engendered by zealous planners indiscriminately using concepts such as "growth pole" and "growth center" is quite considerable still. More often than not, so-called centering policies meant little more than placing investments in cities on the basis of some vaguely defined criterion of "growth potential." Two principles were involved: first, the notion of agglomeration economies which supposedly rendered urban production less costly because producers could tie into a shared system of infrastructure facilities (therefore, the provision of urban services at growth center locations came to be regarded as a major policy variable) and second, the belief—not fully substantiated by empirical research—that economic growth, initially localized in a given center, would, sooner or later, *spread* to surrounding communities.<sup>47</sup> The implied idea that scarce resources would yield larger economic benefits when applied to a limited number of centers having a high potential for economic growth seemed intuitively appealing; also, it enormously simplified the planning problem, since planners could now in good conscience ignore all but a few strategically chosen places; the leftover areas would be taken care of by the spread effects that were presumed to follow.

Aside from the questionable nature of these assumptions, particularly with regard to spread effects, there remained the practical problem of, first, how to identify growth centers and second, how many such centers should be chosen for special treatment.<sup>48</sup>

In the United States, the most eloquent exponent of growth center policies is Niles Hansen, who has advocated their adoption specifically for Appalachia.<sup>49</sup> Outside America, their principal spokesman is Antoni Kuklinski, who, while in charge of regional research at the United Nations Institute for Research in Social Development (UNRISD) in Geneva, sponsored a number of studies concerning the use of growth centers in regional policy. Two collections of these essays have recently been published.<sup>50</sup> An additional influen-

<sup>47</sup>Vida Nichols, "Growth Poles: An Evaluation of Their Propulsive Effect," *Environment and Planning*, vol. 1, no. 2 (1969): 193–208; M. J. Mosley, "The Impact of Growth Centers in Rural Regions. I. An Analysis of Spatial 'Patterns' in Brittany," *Regional Studies*, vol. 7, no. 1 (March 1973): 57–75; Herbert C. Weinand, "Some Spatial Aspects of Economic Development in Nigeria," *The Journal of Developing Areas*, vol. 7, no. 2 (January 1973): 247–264.

<sup>48</sup>Niles M. Hansen, "Criteria for a Growth Centre Policy," in Antoni Kuklinski, ed., *Growth Poles and Growth Centres in Regional Planning* (The Hague: Mouton, 1972), reprinted as chapter 26 in this volume.

<sup>49</sup>Niles M. Hansen, *Intermediate-Size Cities as Growth Centers: Applications for Kentucky, the Piedmont Crescent, the Ozarks, and Texas* (New York: Praeger, 1971).

<sup>50</sup>Antoni Kuklinski and R. Petrella, eds., *Growth Poles and Regional Policies* (The Hague: Mouton, 1972); Antoni Kuklinski, ed., *Growth Poles and Growth Centres in Regional Planning* (The Hague: Mouton, 1972).

tial work is that of E. A. J. Johnson on the role of market centers in rural development.<sup>51</sup>

**Location policy.** How should governments decide on the question of where to encourage industrial development? Outside the literature on growth centers, which is relevant here, relatively little work has been done that would facilitate public choice in this matter. Economic location theory uses a private calculus to determine the best location for an individual firm; geographic location theory (Christaller, Lösch) may provide a basis for judging the merits of alternative *patterns* of location. But neither theory addresses itself to the dilemma of the public policymaker or planner who must find a path between the Scylla of efficiency and the Charybdis of equity in allocating resources to industrial development in particular locations, or who may be asked to select viable industries for promotion in economically backward and declining regions. Both these questions have been addressed in the literature, the first by Alonso,<sup>52</sup> the second by Leo Klaassen.<sup>53</sup> Alonso, who rests his case almost entirely on the single value of maximizing economic growth, resolves the planner's dilemma in favor of efficiency. The problem is put in terms of trade-off: How much growth is the economy prepared to sacrifice for a gain in the distribution of income? In another article Alonso attacks this same question from the standpoint of the individual producer in a developing economy.<sup>54</sup> Not surprisingly, the question of a best location is resolved in substantially similar ways by private and public planners: acting on the basis of an overriding criterion of growth efficiency, both opt for investment in the largest city (or cities) of a country. Alonso further suggests that in the event a public authority wished to alter the location choices of private investors, the size of the subsidy required would probably have to be very large. An economy erected on capitalist premises is therefore likely to increase regional inequalities. Only in the very long run, as Williamson argues (see note 15), will the normal processes of growth in a competitive economy lead to the gradual convergence of regional incomes. (This line of reasoning sets aside the possibility of realizing exceptionally high social returns on capital from either public or private investments in resources development in the "periphery.")

<sup>51</sup>E. A. J. Johnson, *The Organization of Space in the Developing Countries* (Cambridge: Harvard University Press, 1970).

<sup>52</sup>William Alonso, "Urban and Regional Imbalances in Economic Development," *Economic Development and Cultural Change*, vol. 17, no. 1 (October 1968): 1-14, reprinted as chapter 28 in this volume.

<sup>53</sup>Leo H. Klaassen, *Methods of Selecting Industries for Depressed Areas: An Introduction to Feasibility Studies* (Paris: OECD, 1967); idem, *Social Amenities in Area Economic Growth: An Analysis of Methods of Defining Needs* (Paris: OECD, 1968). See also L. M. Mennes, J. Tinbergen, and J. G. Waardenburg, *The Element of Space in Development Planning* (Amsterdam: North Holland Publishing Co., 1969).

<sup>54</sup>William Alonso, "Industrial Location and Regional Policy in Economic Development," Institute of Urban and Regional Development, Working Paper no. 74 (Berkeley: University of California, 1968), reprinted as chapter 3 in this volume.

**Subnational policies for full employment.** Since the original publication of Alonso's article on efficiency vs. equity in regional development, the productive absorption of a rapidly expanding labor force has been elevated to the status of a major development objective. This has been due largely to the efforts of the International Labour Organization.<sup>55</sup> Urban unemployment growing under the impact of rapid rural-to-urban migration is now widely perceived as an urgent social and economic problem in developing societies, at least equal to that of maintaining an accelerated pace of growth in production. The unemployment problem has been studied from a number of different policy perspectives, including that of small-scale industries, intermediate (or progressive) technologies, and rural development.<sup>56</sup> In a paper that elaborates a model of the urban labor market in developing economies, Friedmann and Flora Sullivan<sup>57</sup> propose a far-reaching set of policies that, by fundamentally altering economic relations in the society, would facilitate the adoption and implementation of specific measures aimed at the full development of human potentials, full employment, and the achievements of greater equality in the distribution of income and wealth. The expected impact of these policies on regional development would be pronounced; among other things, they would lend support to the rural market center strategy propounded by Johnson (see note 51) and lead to a regionally more balanced development.

#### Evaluation of Experience

Regional planning has finally come into its own. There are few major countries that have not adopted some form of regional planning. Since 1964, many of these efforts have been the subject of detailed description and critical analysis. There are now more than twenty book-length studies available, covering nearly a score of countries, including a major comparative study covering six countries by Lloyd Rodwin,<sup>58</sup> comparative studies of regional planning in Italy and France (with reference to Great Britain)<sup>59</sup> and in Bolivia, Chile, and Peru,<sup>60</sup> as well as individual studies for the United

<sup>55</sup>International Labour Organization, *The World Employment Programme* (Geneva: ILO, 1971).

<sup>56</sup>Walter Galenson, ed., *Essays on Employment* (Geneva: ILO, 1971); International Labour Organization, *Employment, Incomes, and Equality: A Strategy for Increasing Productive Employment in Kenya* (Geneva: ILO, 1972); Ronald G. Ridker and Harold Lubell, eds., *Employment and Unemployment Problems of the Near East and South Asia*, 2 vols. (Delhi: Vikas Publications, 1971).

<sup>57</sup>John Friedmann and Flora Sullivan, "The Absorption of Labor in the Urban Economy: The Case of Developing Countries," *Economic Development and Cultural Change*, April 1974, reprinted as chapter 21 in this volume.

<sup>58</sup>Lloyd Rodwin, *Nations and Cities: A Comparison of Strategies for Urban Growth* (Boston: Houghton Mifflin Co., 1970).

<sup>59</sup>Kevin Allen and M. C. MacLennan, *Regional Problems and Policies in Italy and France* (Beverly Hills: Sage Publications, 1970).

<sup>60</sup>Sergio Boisier, *Polos de desarrollo: hipotesis y políticas: estudio de Bolivia, Chile y Perú* (Geneva: United Nations Research Institute for Social Development, January 1972).

States,<sup>61</sup> Canada,<sup>62</sup> France,<sup>63</sup> England,<sup>64</sup> Ireland,<sup>65</sup> Japan,<sup>66</sup> Israel,<sup>67</sup> Mexico,<sup>68</sup> Venezuela,<sup>69</sup> Turkey,<sup>70</sup> South America,<sup>71</sup> Southeast Asia,<sup>72</sup> Eastern Europe,<sup>73</sup> and Yugoslavia.<sup>74</sup> In addition, we have a detailed case study of planning for the resource frontier region of the Venezuelan Guayana<sup>75</sup> and the proceedings of an international conference on experience with policies for economically backward regions.<sup>76</sup>

Despite this wealth of information, it cannot be claimed that we really know whether and in what sense regional policy and planning may be said to work. There can be little doubt, however, that such planning corresponds to a variety of perceived national needs (see following section). Also, a good deal else may be learned from these accounts. One attempt to draw appropriate lessons from experience is Friedmann's assessment of the principal methods used in implementing regional policies.<sup>77</sup> An interesting and even somewhat

<sup>61</sup>Niles M. Hansen, *Rural Development and the Urban Crisis: A Strategy for Regional Development* (Bloomington: Indiana University Press, 1970); Gordon C. Cameron, *Regional Economic Development: The Federal Role* (Washington: Resources for the Future, 1970); John H. Cumberland, *Regional Development Experiences and Prospects in the United States of America* (The Hague: Mouton, 1971).

<sup>62</sup>T. N. Brewis, *Regional Economic Policies in Canada* (Toronto: Macmillan Co. of Canada, 1969).

<sup>63</sup>Niles M. Hansen, *French Regional Planning* (Bloomington: Indiana University Press, 1968).

<sup>64</sup>Gavin McCrone, *Regional Policy in Britain* (London: George Allen and Unwin, 1969).

<sup>65</sup>Helen B. O'Neill, *Spatial Planning in the Small Economy: The Case of Ireland* (New York: Praeger, 1971).

<sup>66</sup>Masahiko Honjo, *Trends in Japanese Development Planning* (Tokyo University, January 1971); Ryoichi Kabaya, "Development of Poor Regions: General Considerations and the Case of Japan," Institute of Urban and Regional Development, Working Paper no. 159 (Berkeley: University of California, October 1971).

<sup>67</sup>Eliezer Brutzkus, *Regional Policy in Israel* (Jerusalem: Ministry of Interior, Town and Country Planning Department, 1970).

<sup>68</sup>David Barkin and Timothy King, *Regional Economic Development: The River Basin Approach in Mexico* (Cambridge: University Press, 1971).

<sup>69</sup>John Friedmann, *Regional Development Policy: A Case Study of Venezuela* (Cambridge: MIT Press, 1966); see also idem, *From Doctrine to Dialogue* (Syracuse: Syracuse University Press, 1965).

<sup>70</sup>Malcolm D. Rivkin, *Area Development for National Growth: The Turkish Precedent* (New York: Praeger, 1965).

<sup>71</sup>Walter B. Stöhr, *El desarrollo regional en America Latina: experiencias y perspectivas* (Buenos Aires: Ediciones SIAP, 1972). (An English version of this book is forthcoming from Mouton.)

<sup>72</sup>Louis Lefebvre and Mrinal Datta-Chaudhuri, *Regional Development Experiences and Prospects in South and Southeast Asia* (The Hague: Mouton, 1971).

<sup>73</sup>Kosta Mihailovic, *Regional Development Experiences and Prospects in Eastern Europe* (The Hague: Mouton, 1972).

<sup>74</sup>Jack C. Fisher, *Yugoslavia, A Multi-national State: Regional Differences and Administrative Response* (San Francisco: Chandler, 1966).

<sup>75</sup>Lloyd Rodwin and Associates, *Planning Urban Growth and Regional Development: The Experience of the Guayana Program of Venezuela* (Cambridge: MIT Press, 1969).

<sup>76</sup>E. A. G. Robinson, ed., *Backward Areas in Advanced Countries: Proceedings of a Conference Held by the International Economic Association* (New York: St. Martin's Press, 1969).

<sup>77</sup>Friedmann, *Urbanization, Planning, and National Development*, chapter 12, "The Implementation of Regional Development Policies: Lessons of Experience."

embarrassing question may be asked in this connection. To what extent have the theories discussed earlier in this chapter influenced the actual practice of planning?

The answer is, probably, not much.<sup>78</sup> Still, theoretical and practical knowledge have not been wholly unrelated. Over the last decade, for instance, theorists have argued a systems approach to regional planning;<sup>79</sup> in practical terms, this has meant coordinating regional policies at the national level. Or, to cite another example, central place theory, which plays an important role in spatial organization, has found its practical counterpart in growth pole and centering policies which are now widely applied. Regardless of the practical success of these policies, and they are still in dispute, attention to subnational development has provided central planners with a treasure hoard of concrete information about social, economic, and physical conditions in different parts of the country;<sup>80</sup> in many cases, this has brought into being an administrative apparatus to assist in decision-making and coordination of programs tailored for regional needs (for details, see Walter Stöhr's study of regional planning in South America, cited in note 71).

### The Tasks of Regional Planning

Regional planning is fundamentally concerned with the *where* of economic development.\* But why is this question considered important in so many different country settings? Do spatial arrangements make a difference for anything that might be expressed in terms of some broad welfare objectives?

The general answer to these questions is that regional planning strives to achieve a better integration of spatially organized economies on a basis of *interdependence (and reciprocity)* rather than *dependence (and exploitation)*. This general purpose of regional planning, however, encompasses a number of concrete goals and objectives that tend to sort themselves out according to whether the concern is with industrially advanced or developing countries.

In the former, the emphasis tends to be on the following kinds of problems,

\*To state this as the basic question is not in any way intended to denigrate the equally important questions of *what* activities to select and *how* and *by what means* to carry out a program of regional development. The spatial framework for action is essentially a coordinative one. The regional planner who is trained to look at the interconnections among activity patterns and to be concerned with settlement processes, such as migration, that are intimately related to differential economic growth, must therefore work closely with sectoral specialists in the preparation of subnational development programs. On the other hand, he will be the person principally responsible at the level of national planning for working out the policies and criteria which are meant to guide both the investment and fiscal policies of government in their spatial dimensions.

<sup>78</sup>Friedmann, *Urbanization, Planning, and National Development*, chapter 13, "The Uses of Regional Science: Policy Planning in Chile."

<sup>79</sup>J. Brian McLoughlin, *Urban and Regional Planning: A Systems Approach* (London: Faber & Faber, 1969).

<sup>80</sup>Torsten Hägerstrand and Antoni Kuklinski, eds., *Information Systems for Regional Development: A Seminar*, Lund Studies in Geography, Ser. B, Human Geography, no. 37 (Lund: C. W. K. Gleerup, 1971).



among others, which are responsive to spatially oriented policies:

- economically backward or depressed regions
- economically declining or obsolescing regions
- inequalities in the indicators of social and economic development among culturally defined or political regions
- spatial reorganization of settlement and economic activity patterns necessitated by internal migration, changes in major transport technologies, common market policies, etc.
- accelerated outward movement of central city populations into "urban fields"<sup>81</sup>
- major environmental problems associated with large-scale and densely populated urban settlements.

In the developing countries, the emphasis tends to be on another set of issues, including:

- regional coordination of national investment policies and programs
- excessive centralization in the territorial organization of power
- growing urban primacy and underutilization of natural resources on the periphery of major core regions
- underutilization of manpower and human talent
- the incorporation of internally dominated peripheral regions into a single national economic system on a basis of greater equality in living levels and reciprocity in exchange
- dualistic development of rural and urban economies
- destructive use of land resources.

These recurring problem sets are closely interrelated, and their "solution" demands a national approach. The United States, which, objectively speaking, has many of these problems in common with other industrially advanced societies, nevertheless has shown relatively little interest in regional policies and planning. Except for occasional, halfhearted attempts to deal with regional obsolescence, economic decline, and national settlement patterns, the United States has probably shown less commitment to regional policies and planning than any comparable nation. Attempts to provide a general explanation for this phenomenon are bound to be risky. A few reasons, however, may be cited, even though it will be difficult to assign to each their relative importance.

Unlike many countries in Europe and elsewhere, the United States has no significant and relatively self-contained cultural regions. On the whole, the American people move easily within a single national space, undivided by

<sup>81</sup>For the concept of "urban field," see note 10 and John Friedmann, "The Future of the Urban Habitat," in Donald MacAllister, ed., *A New Focus for Land Use Planning* (Washington: National Science Foundation Press, 1973), pp. 96-134.

major differences in language, religion, and general culture. Some older regions, such as New England and the traditional South, resemble in some respects the more tightly circumscribed historical regions of Europe and have, indeed, received more attention from regional planners than other parts of the country; on balance, however, the practical effects of this have been negligible. To a surprising extent, Americans have valued *place* less than geographic *mobility*. In Hirschman's language, they have more frequently exercised their right to exit than their constitutional right to voice.<sup>82</sup>

This peculiar behavior can probably be traced to an abiding faith in the ideology of individualism in which the aggressive pursuit of self-interest is legitimized at the expense of a sense of communal responsibility. Americans have shown remarkably little concern for the shared cultural values of the past. The social costs of so-called progress, while frequently regretted, are, by and large, accepted as inevitable. In any event, a functional articulation of interests is more strongly approved than the articulation of regional interests. Regional inequalities in economic well-being—no less than other inequalities—are believed to be the “natural” result of an economic system supportive of individual efforts at bettering oneself. The system has never been seriously challenged except by small, dissident minorities, with little support among the mass of the population. Moreover, the system is, in any event, believed to be gradually moving toward greater regional equality of outcomes by permitting the free flow of labor, capital, commodities, and ideas. Any restraint placed on these flows by public action would not only be thought to run counter to the traditional American beliefs in a man's right to do as he pleases but also lead to an inefficient allocation of resources and thus to a reduction in the growth of the national product.<sup>83</sup>

Finally, Americans seem to accept as axiomatic a belief in the goodness of political pluralism and the design of governmental arrangements according to a principle of checks and balances. A system so devised is antithetical to the idea of area-wide, systemic planning. Whether a growing awareness of the finiteness of natural resources and declining environmental quality will introduce significant changes in these fundamental attitudes and beliefs remains to be seen, but it is unlikely.

### The Next Decade

Regional planning has drawn nourishment from two major intellectual traditions: economics and geography. From the first, planners learned to appreciate the virtues of marginal decision analysis (particularly in the context of economic location theory), the powers of deductive logic, and the beauties of spatial equilibria; from the second, they learned to apply systems analysis and inductive reasoning in the study of spatial behavior. As economics emphasized the mobility of labor and capital, so geography laid stress on the

<sup>82</sup>Albert Hirschman, *Exit, Voice, and Loyalty* (Cambridge: Harvard University Press, 1970).

<sup>83</sup>George H. Borts and Jerome L. Stein, *Economic Growth in a Free Market* (New York: Columbia University Press, 1964).

structural characteristics of particular places. It is only natural, therefore, that these two traditions should eventually have been joined, a feat that was accomplished with the creation of the synthetic field of regional science. In fact, the publication of this volume celebrates the twentieth anniversary of the International Regional Science Association, founded by Walter Isard in 1954.

The planner's craft builds upon experience as well as theory. It is hoped that both will be advanced during the coming decade. With regard to the first, there is now a pressing need, particularly in the industrially advanced societies, to relate questions of location and spatial organization to policies for environmental protection and social amenities.<sup>84</sup> In the newly industrializing countries, on the other hand, growing shortages of food combined with rising urban unemployment point to the urgency of turning attention to problems of comprehensive rural development.<sup>85</sup> Both types of experience will, in due course, yield studies from which much may be learned. With regard to theory, however, a more systematic approach is indicated. In concluding this chapter, therefore, I should like to suggest what I believe to be an appropriate research agenda for regional planning during the years ahead.<sup>86</sup>

**The effects of imbalanced power relationships on regional growth and development.** The unequal distribution of economic and political power is one of the principal determinants of observed patterns of regional economic growth and settlement. Yet relatively little is known of the mechanisms through which these imbalances affect investment and location behavior or, for that matter, how the politics of interregional conflict influence the spatial allocations of major power centers. These aspects of power and power conflict deserve closer study if we are to gain a better understanding of the dynamics of regional development in culturally heterogeneous as well as homogeneous societies, in federal as well as unitary governmental systems. Of particular interest is the way in which different forms of external dependency influence the evolution of spatial patterns of peripheral economies and the methods through which this influence, intended or unintended, is being asserted.

**The spatial effects of accelerated urbanization in developing countries.** The spatial impact of urbanization can be subsumed under the broad categories of *spread* and *backwash* effects. We need to increase the number of studies that attempt to show, in a temporal and spatial framework, the extent of these effects as they are superimposed upon each other and the conditions under which one or the other set of effects tends to predominate. Studies need to be

<sup>84</sup>MacAllister, *A New Focus*.

<sup>85</sup>Raanan Weitz, *From Peasant to Farmer: A Revolutionary Strategy for Development*, A Twentieth Century Fund Study (New York: Columbia University Press, 1971). See also H. I. Ajaegbu, "Regional Planning and the Rural Areas of Nigeria," in K. M. Barber, ed., *Planning for Nigeria* (Ibadan: Ibadan University Press, 1972), chapter 7.

<sup>86</sup>For an earlier discussion of possible research topics, see John Friedmann, "Regional Planning and Nation-Building: An Agenda for International Research," *Economic Development and Cultural Change*, vol. 16, no. 1 (October 1967): 119-130.

carried out for growth centers of different size and subject to varying rates of growth, in urban systems having different forms of hierarchical structure, and under different agricultural systems of land tenure, production, and marketing.

**The dynamic relationship between agricultural production and urban growth.** The only aspect of this relationship which has thus far received attention is rural-to-urban migration. To what extent can migration be slowed down by improving income earning opportunities and services in rural areas? What role should be assigned to rural marketing centers in improving the access of rural populations to essential services and in providing opportunities for nonfarm employment? Should a strategic role in rural development be assigned to agricultural processing industries and, if so, what results can be expected from this approach? What changes in agricultural marketing organization should be introduced simultaneously to benefit producers and to supply urban populations at the lowest possible cost? What role might be assigned to cities as staging areas for the agricultural settlement of new frontier areas, especially in the tropics?

Answers to these questions are becoming increasingly urgent. On the other hand, in light of growing unemployment and underemployment in cities, the functioning of urban labor markets in the developing countries needs to be better understood. What is the structure of these markets, and how are various submarkets interconnected? What is the extent of vertical and horizontal job mobility, and how is this related to the characteristics of the urban labor force? How much information concerning employment conditions in cities filters back into the rural areas, and how is this information evaluated with respect to the decision to migrate? What evidence is there that the level of urban unemployment acts as a regulator for migration from rural areas? And finally, what are the marginal and average costs of absorbing population into cities of different sizes? While these questions by no means exhaust the general topic, they are indicative of the extent of our present ignorance.

**The structure and evolution of urban fields in postindustrial societies and their impact on the natural environment.** In the highly advanced industrial and post-industrial societies, but especially in the United States, so-called urban fields are replacing the traditional city as the basic ecological unit of urban life. These fields are defined at their edges by the use made of the land for recreation and second-home communities, but increasingly also by the location of higher-level services such as insurance offices, medical centers, research laboratories, and educational institutions. This new form of the emerging postindustrial "city," made possible by the combined influence of high family incomes, the ubiquitous use of the automobile, and advanced communication and computer technologies, has scarcely begun to be studied, so that opportunities for significant research abound. The topics requiring exploration include changing land use and circulation patterns, population densities and land use patterns related to daily, monthly, and seasonal cycles,

the impact of these patterns on the availabilities of open space and scenic amenities, problems of high-density use of environmental resources, "invisible" design for high-density recreation areas, emerging activity clusters and their functional differentiation, evolving political attitudes, motivations, and loyalties, the social correlates of emerging physical patterns, appropriate organizational forms for carrying out essential governmental functions at the scale of urban fields, fiscal problems and policies, effects of an expanded scale of urban living on the older centers within the region, and problems of adapting these centers to the emerging forms of settlement and use.

**The evaluation of regional planning practice and the effectiveness of implementing procedures.** Despite the growing number of case studies of regional planning, we are still lacking a systematic evaluation of the effectiveness of implementing mechanisms for regional policies. Such studies are best undertaken on a comparative basis, with due regard to varying institutional settings and, in particular, should be extended to include the experience of socialist countries. The "tools" for implementing regional development and for affecting the spatial patterns of settlement, production, and welfare are limited in number and include location (or capital) subsidies, labor (or wage) subsidies, other fiscal policies related to taxation and import tariffs, direct controls over migration, government investments in infrastructure and directly productive activities, new towns, land use controls, licensing arrangements, administrative decentralization, increasing regional autonomy in matters affecting development, and central planning and budgetary controls. One would not expect each item in this repertory of potential actions to be equally suitable for every purpose nor, indeed, to be effective under any set of institutional arrangements. The detailed and comparative analysis of these methods, both in theory and application, must consequently rank high on any agenda for research whose aim is to improve the efficacy of regional planning.